UNIVERSITATEA DIN BUCUREŞTI

FACULTATEA DE PSIHOLOGIE ŞI ŞTIINȚELE EDUCAȚIEI

DEPARTAMENTUL DE ÎNVĂȚĂMÂNT LA DISTANȚĂ

STATISTICĂ PSIHOLOGICĂ ŞI PRELUCRAREA COMPUTERIZATĂ A DATELOR

Noțiuni statistice fundamentale
Statistici descriptive

Notă:

Conținutul modulelor este suficient de dezvoltat pentru însuşirea cursului, dar versiunea completă a materiei de curs se găseşte în volumul: **M. Popa, Statistică pentru psihologie. Teorie şi aplicații SPSS, apărut la Editura Polirom (2008).**

Conf. univ. dr. Marian Popa
e-mail: popamarian@gmail.com
web page: www.mpopa.ro

Universitatea din Bucureşti
Editura CREDIS
2008

Copyright © DEPARTAMENT ID 2008

Acest material este destinat uzului studenţilor Universităţii din Bucureşti, forma de învăţământ la distanţă.
Conţinutul cursului este proprietatea intelectuală a autorului/autorilor; designul, machetarea şi transpunerea în format electronic aparţin Departamentului de Învăţământ la Distanţă al Universităţii din Bucureşti.

Universitatea din Bucureşti
Editura CREDIS
Bd. Mihail Kogălniceanu, Nr. 36-46, Corp C, Etaj I, Sector 5
Tel: (021) 315 80 95; (021) 311 09 37, 031 405 79 40, 0723 27 33 47
Fax: (021) 315 80 96
Email: credis@credis.ro
Http://www.credis.ro

Copyright © DEPARTAMENT ID 2008

Informații cu privire la organizarea şi desfăşurarea cursului

- În conformitate cu specificul învățământului la distanță, cursul se bazează pe studiu individual şi activități tutoriale.
- **Suportul de curs** pentru studiul individual este constituit din:
  - **modulele sintetice** distribuite la începutul fiecărui semestru;
  - **manualul cursului: "**Marian Popa, 2008, Statistică pentru psihologie. Teorie şi aplicații SPSS, Editura Polirom". Poate fi achiziționat din librării sau poate fi consultat la biblioteca facultății.
  - **materiale suplimentare** care vor fi postate pe pagina web www.mpopa.ro
- **Activitățile tutoriale** sunt planificate, de regulă, lunar, şi au un caracter interactiv. Studenții pot solicita explicații sau pot pune întrebări în legătură cu tematica tutorialului. Pe durata semestrului se pot solicita explicații suplimentare cu privire la materia de curs, pe cale electronică, de la titularul de curs sau de la tutore.
- **Lucrările de control** vor fi asociate fiecărui modul şi vor fi transmise, de preferință, sub formă electronică. Pentru a fi acceptată, o lucrare de control trebuie să fie transmisă în timpul limită fixat, să denote o însuşire suficientă a materiei şi efortul personal pentru realizarea ei. Lucrările nu vor fi evaluate cu note.
- **Evaluarea** se face pe bază de examen scris şi constă dintr-un număr de întrebări punctuale, care cer un răspuns scurt şi la obiect. Înainte de examen se va transmite un set de întrebări orientative, din genul celor de la examen. Nota de examen va fi corectată în funcție de lucrările de control acceptate astfel:
  - pentru trei lucrări acceptate nu se scade nici un punct
  - pentru două lucrări acceptate se scade un punct
  - pentru o lucrare acceptată se scad două puncte
  - pentru nicio lucrare acceptată se scad trei puncte
- **Condiții de echivalare a examenului**. Studenții/studentele care au urmat un curs de statistică de nivel universitar (cursuri de licență) pot solicita echivalarea examenului în următoarele condiții:
  - programa echivalentă (dovada de face prin extras după tematica de curs)
  - nota obținută să fie cel puțin 7 (nu se echivalează notele de 5 şi 6)
  - cererea de echivalare se va face cel mai târziu până la data primului tutorial, după acest moment nicio cerere nu va mai fi luată în considerare.
- **Procedura de echivalare**
  - Cerere adresată decanului facultății, avizată de titularul de curs, şi o copie a foii matricole

Copyright © DEPARTAMENT ID 2008

CUPRINS

NOȚIUNI STATISTICE FUNDAMENTALE 5
Obiectivele unității de învățare 5
Definiția şi rolul statisticii în psihologie 5
Utilitatea statisticii în practica psihologică 6
Măsurarea în psihologie 6
Scala nominală 7
Scala ordinală 7
Scala de interval 8
Scala de raport 8
Sarcina de lucru nr. 1. 1 9
Concepte statistice fundamentale 9
Noțiunea de variabilă statistică 9
Variabile dependente şi variabile independente 9
Sarcină de lucru nr. 1. 2 10
Variabile continue şi variabile discrete 10
Populație şi eşantion 10
Sarcină de lucru nr. 1. 3 11
Statistica descriptivă şi statistica inferențială 12
Statistica parametrică şi statistica neparametrică 12
Studii experimentale şi studii observaționale 12
Rezumatul unității de învățare 13
Răspunsuri corecte la sarcinile de lucru 14
Lucrarea de evaluare nr. 1.1 15
Bibliografie 15
STATISTICI DESCRIPTIVE 16
Obiective de învățare şi informații introductive 16
Statistici descriptive globale 17
Analiza de frecvențe 17
Analiza de frecvențe simple 17
Analiza de frecvențe grupate 19
Sarcina de lucru nr. 2.1 21
Reprezentarea grafică a datelor 22
Graficul de tip bară 22
Histograma 23
Poligonul de frecvențe 23
Graficul frecvenței cumulate 24
Graficul circular 24
Reprezentarea de tip *stem-and-leaf* (*stem plot*) 25
Stem-and-Leaf 1
Sarcina de lucru nr. 2.2 26
Indicatori statistici descriptivi 26
Indicatori ai tendinței centrale 26
Modul (Mo) 26
Mediana (Me) 27
Media aritmetică (m) 27
Sarcina de lucru nr. 2.3 29
Indicatori ai împrăştierii 29
Amplitudinea absolută (R de la Range) 30
Amplitudinea relativă 30

Copyright © DEPARTAMENT ID 2008

Abaterea quartilă (cvartilă, intercvartilă) (R_Q) 31
Abaterea semi-interquartilă (R_{SQ}): 31
Abaterea medie (d de la deviație medie) 32
Dispersia (varianța, abaterea medie pătratică) 33
Abaterea standard 33
Coeficientul de variație 35
Indicatori ai formei distribuției 36
Sarcina de lucru nr. 2.4 38
Valori extreme ale distribuției 39
Tratarea valorilor extreme 40
Rezumatul unității de învățare 41
Răspunsuri corecte la sarcinile de lucru 41
Lucrarea de evaluare nr. 1.2 42
Bibliografie minimală 43

Copyright © DEPARTAMENT ID 2008

NOȚIUNI STATISTICE FUNDAMENTALE

Obiectivele unității de învățare

Parcurgerea acestei unități, va permite studenților:

- ✓ să explice utilitatea analizei statistice în domeniul psihologiei
- ✓ să definească noțiunea de variabilă statistică
- ✓ să identifice diferite tipuri de variabile statistice
- ✓ să definească noțiunile de eşantion şi de populație statistică
- ✓ să explice specificul statisticii descriptive şi inferențiale
- ✓ să explice diferența dintre statistica parametrică şi neparametrică
- ✓ să identifice scalele de măsurare ale variabilelor statistice

Definiția şi rolul statisticii în psihologie

Definiție:

Statistica psihologică este disciplina care se ocupă cu analiza datelor care descriu aspecte de natură psihică, individuală sau colectivă, în scopul de a le prezenta sintetic, sub formă numerică sau grafică, de a le analiza şi de a extrage concluzii pe seama lor.

Faptele de natură psihică sau cu semnificații psihologice care fac obiectul măsurării şi al analizei statistice, pot fi extrem de variate: genul (masculin/feminin), inteligența, timpul de reacție, atitudinile, nivelul motivației, nivelul diverselor caracteristici psihice (sociabilitate, anxietate, emotivitate)

Metoda ştiințifică

Statistica este un instrument al metodei ştiințifice în psihologie

În esență, ştiința este o metodă, un mod specific de a afla răspunsuri la întrebările pe care ni le punem. Principalele ei caracteristici sunt: căutarea unor reguli generale (legități), colectarea unor dovezi obiective, operarea cu afirmații controlabile, atitudine sceptică față de cunoştințele acumulate, atitudine deschisă față de orice informații noi, creativitate şi transparență.

În mod practic, un demers de tip ştiințific porneşte de la identificarea unei probleme, urmată de o serie standardizată de etape de găsire a răspunsului adecvat: generarea unei ipoteze în legătură cu răspunsul posibil, testarea ipotezei (prin experiment sau altă metodă empirică), analizarea datelor recoltate, emiterea unei decizii de confirmare sau de infirmare a ipotezei. În faza următoare, ipoteza poate fi rafinată iar procesul se reia atâta timp cât problema prezintă un interes de cunoaştere. În acest proces, statistica se ocupă cu tratarea datelor numerice prin sintetizarea lor, într-o primă fază, şi prin proceduri de analiză care să fundamenteze o

Copyright © DEPARTAMENT ID 2008

decizie lipsită de subiectivitate cu privire la adevărul ipotezei, în a doua fază.

Exemplu: Un psiholog observă, printre cunoscuții săi, că cei care fumează sunt, de regulă, mai emotivi decât cei care nu fumează. Dorind să verifice dacă acest lucru este adevărat şi nu doar o simplă impresie, selecționează două grupuri de persoane, fumători şi nefumători, cărora le aplică un instrument de evaluare a trăsăturilor de personalitate, printre care şi manifestările de tip afectiv. Rezultatele individuale se concretizează în scoruri numerice pentru fiecare trăsătură. Utilizând proceduri statistice, calculează media scorului pentru fiecare dintre trăsăturile investigate. Să zicem că, în final, constată că fumătorii au obținut un scor la anxietate de 14.3, comparativ cu nefumătorii, care au obținut un scor mediu de 12.7. Este această diferență suficientă pentru a concluziona că fumătorii au un nivel mai ridicat de anxietate, sau diferența obținută nu este decât una neimportantă, care a apărut întâmplător la grupurile investigate şi nu poate nu poate fi generalizată dincolo de acestea?

Pentru a răspunde la această întrebare trebuie utilizată o anumită procedură statistică, care va face obiectul unei teme de curs.

Utilitatea statisticii în practica psihologică

Pentru că este dificil să înveți ceva fără a avea o imagine clară a utilității acelor cunoştințe, iată câteva argumente în sprijinul ideii că utilizarea statisticii face parte integrantă din activitatea curentă a unui psiholog:

- ✓ Elaborarea şi utilizarea testelor psihologice
- ✓ Selecția psihologică
- ✓ Studii şi cercetări psihologice: identificarea caracteristicilor unor categorii de persoane (de ex., diferențe dintre bărbați şi femei, dintre diferite metode de terapie etc.)

Statistica este un instrument indispensabil în practica profesională a psihologului

Statistica oricât de sofisticate ar fi, nu dă psihologiei, prin ea însăşi, un caracter de ştiință. Ştiința este o metodă, un model de cunoaştere a realității, o cale prin care se explorează necunoscutul şi se fac previziuni. Statistica, la fel ca şi metodele psihologice, nu sunt decât instrumente utile, indispensabile, pentru abordarea ştiințifică a fenomenelor psihice.

Măsurarea în psihologie

În esență, a măsura înseamnă a atribui numere sau simboluri unor caracteristici ale realității obiective sau subiective, în funcție de anumite aspecte cantitative sau calitative care le caracterizează. În acest mod relația dintre numere sau simboluri ajunge să reflecte relația dintre caracteristicile cărora le-au fost atribuite. Modul în care sunt atribuite numere sau simboluri pentru a măsura ceva, se numeşte „scală de măsurare”.

A măsura în înseamnă a atribui numere sau simboluri unor caracteristici ale realității.

Copyright © DEPARTAMENT ID 2008

Statistica operează cu valori numerice sau de altă natură, care rezultă dintr-un proces de măsurare. Dar numerele, deşi au aceeaşi formă, nu sunt asemănătoare unele cu altele. Ele pot avea diferite semnificații sau proprietăți în funcție de tipul de măsurare din care rezultă. În funcție de cantitatea de informație pe care o reprezintă valorile, ca rezultat al procesului de măsurare, putem distinge mai multe tipuri de scale de măsurare:

Scala nominală

Valorile de tip nominal, „denumesc" indivizi sau categorii de indivizi.

O măsurare pe scală nominală înseamnă, de fapt, a plasa obiectele în diferite clase. În acest caz, o valoare nu este cu nimic mai mare sau mică decât altă valoare. Un exemplu la îndemână este „valoarea" atribuită genului. Ea poate fi codificată cu „M" sau „F", ori, la fel de bine cu „2" sau „1". În acest caz, respectivele „valori" nu sunt decât simboluri ale unei anumite calități pe care o ia caracteristică de gen a unei persoane. Cu alte cuvinte, într-un asemenea caz „2" nu înseamnă că este „mai mult" sau „mai bun" decât „1", ci doar faptul că este „diferit" de acesta. Vom observa că ambele codificări de mai sus sunt arbitrare, în locul lor putând utiliza orice alte simboluri, pe bază de convenție.

Variabilele măsurate pe scale de tip nominal pun în evidență diferențe calitative între valori. Alte exemple de variabile exprimate pe scale nominale: bolile psihice (paranoia, depresie, nevroză), tipurile temperamentale (sanguin, coleric, flegmatic, melancolic), specialitatea universitară (psihologie, chimie, matematica), lateralitatea (dreptaci, stângaci), religia (ortodox, catolic).

Valorile de tip nominal pot fi, la rândul lor, de două feluri:

- De identificare, atunci când o valoare are rolul de codificarea identității, referindu-se în mod unic la o anumită persoană (de ex., codul numeric personal, sau un număr de identificare în cadrul unui experiment psihologic).
- Categoriale, atunci când desemnează forme pe care le ia o variabilă (tipul de liceu absolvit: „teoretic", „industrial", „artistic"; tipurile temperamentale: „sanguin", „coleric", „flegmatic", „melancolic"). Această formă este în mod obişnuit întrebuințată în psihologie, ori de câte ori este necesară repartizarea subiecților în diverse clase sau categorii, în funcție de prezența sau absența anumitor caracteristici.

Valorile măsurate pe o scală de tip nominal au un caracter calitativ şi nu suportă operații numerice, altele decât cele de sumarizare (numărare, procente).

Scala ordinală

Valorile de tip ordinal exprimă poziția, rangul, unei valori dintr-o serie de valori.

Valorile plasate pe o scală de tip ordinal au o anumită semnificație cantitativă. O anumită valoare este "mai mare" sau "mai bună" decât alta, aflată sub ea. Implicit, ea poate fi "mai mică" sau mai "puțin bună" decât altă valoare, aflată deasupra ei. Dacă o anumită persoană este mai preferată decât alta şi atribuim celei primei valoarea 1, iar celei de-a doua valoarea 2, atunci cele două valori se exprimă pe o scală de tip ordinal, care indică doar ordinea preferinței şi nu măsura intensității acestei preferințe.

Exemple: ordinea de rang la nivelul unei clase, în funcție de notele şcolare, ordinea copiilor la naştere.

Copyright © DEPARTAMENT ID 2008

Variabilele ordinale pot fi şi ele de tip categorial, atunci când grupurile definite de valorile variabilei pot fi aranjate într-o ordine naturală. De exemplu: valorile asociate vârstei astfel: „1”=20-30 de ani, „2”=31-40 de ani, „3”=41-50 de ani, sau apartenența la o anumită categorie valorică, rezultată prin evaluarea la un examen cu calificative (foarte bun, bun, mediu, rău, foarte rău).

Scala de interval

O variabilă măsurată pe o scală de interval ne oferă informații nu doar despre ordinea de mărime, ci şi despre „dimensiunea” exactă a caracteristicii măsurate. Valorile de acest tip au un caracter cantitativ, exprimat numeric, iar intervalele dintre ele sunt egale.

Valorile de tip interval exprimă mărimea, cantitatea, în raport cu alte valori.

Exemple:

- temperatura, măsurată pe o scală Celsius. Dacă într-o zi se măsoară 5 grade iar în ziua următoare 10 grade, se poate spune cu precizie că a doua zi a fost cu 5 grade mai cald;
- coeficientul de inteligență măsurat, să zicem, prin numărul de răspunsuri corecte la un test. În acest caz, un rezultat de 30 de răspunsuri corecte este cu 10 unități mai mare decât 20 sau cu 5 unități mai mic decât 35;
- scorurile la testele de personalitate.

Ceea ce este caracteristic valorilor măsurate pe scală de interval este absența unei valori zero absolute, adică absența totală a caracteristicii măsurate. În consecință, valorile de acest tip nu ne permit evaluări de genul: „*O temperatură de 10 grade Celsius este de două ori mai mare decât una de 5 grade Celsius*” sau, „*O persoană care a obținut un scor de 30 de puncte este de două ori mai inteligentă decât una care a obținut 15 puncte*”. Aceasta, deoarece nici temperaturile măsurate pe scala Celsius şi nici inteligența, nu au o valoare 0 absolută (dacă acceptăm că nici un om viu nu are inteligență nulă).

Scala de raport

Valorile exprimate pe o scală de raport dețin cel mai înalt grad de măsurare. Pe lângă egalitatea intervalelor, specifică scalei de interval, acest tip de valori se raportează şi la o valoare 0 absolut (nu este posibilă nici o valoare mai mică de 0). Din acest motiv, este permisă aprecierea raportului dintre două valori.

Valorile de tip raport exprimă mărimea, cantitatea, în raport cu alte valori dar şi cu o valoare absolută.

Exemple

- dacă ne referim la temperaturi, atunci scala Kelvin, este un bun exemplu (0 Kelvin este temperatura minimă absolută)
- timpul
- numărul de răspunsuri corecte sau de erori, la un test psihologic

În psihologie puține sunt variabilele acceptate ca fiind măsurate pe scala de raport, deoarece sunt puține situațiile în care avem de a face cu caracteristici ce pot lua valoarea 0 absolut.

La fel ca şi valorile măsurate pe scale de interval, valorile măsurate pe scală de raport suportă toate transformările matematice posibile. Din acest motiv, în practică, valorile măsurate pe scală de interval sau de

Copyright © DEPARTAMENT ID 2008

raport sunt considerate similare, fiind prelucrate prin acelaşi gen de proceduri statistice. Ca urmare, în acest caz, se spune că o variabilă este măsurată pe o „scală de interval/raport".

Temă de reflecție:
Dacă evaluăm scalele în funcție de nivelul de măsurare ale fiecăreia, pe ce scală de măsurare putem considera că se plasează?
.....

Sarcina de lucru nr. 1. 1
Identificați natura scalei de măsurare pentru următoarele variabile.
Scrieți răspunsul şi apoi verificați corectitudinea la pagina 21

		Tipul scalei
1	Apartenența la o anumită minoritate etnică, codificată astfel: 1. lipoveni; 2. români; 3. polonezi; 4. maghiari; 5. italieni; 6. armeni	
2	Latența reacției la un stimul auditiv, măsurată în sutimi de secundă	
3	Atitudinea față de statistică măsurată pe o scală continuă de la 1 (absolut antipatică) la 10 (absolut simpatică)	
4	Numărul de răspunsuri corecte la un test de calcule aritmetice	
5	Poziția pe o listă la un concurs de admitere organizată în ordinea mediei	

Concepte statistice fundamentale

Noțiunea de variabilă statistică

Statistica se ocupă cu studiul variabilelor, adică al variabilității umane.

Înțelegem prin variabilă statistică o caracteristică a realității care poate lua valori diferite de la persoană la persoană sau în situații diferite. De exemplu, un cercetător doreşte să verifice ipoteza că persoanele care beau cafea seara, adorm mai greu decât cele care nu beau. În acest caz, avem de a face cu două variabile statistice: timpul de adormire, care ia poate fi măsurat în minute, şi consumul de cafea, care este „prezent" la unele persoane şi „absent" la altele. Dacă latența somnului ar fi aceeaşi la toți oamenii, indiferent de condiții sau situații, atunci aceasta nu ar mai fi o variabilă ci o constantă şi nu ar mai prezenta interes pentru analiză statistică.

Variabile dependente şi variabile independente

Statistica se ocupă cu studiul relațiilor dintre variabile

În esență, un studiu statistic îşi propune evidențierea legăturilor dintre diverse caracteristici ale realității (variabile). În acest context, există variabile ale căror valori sunt dependente, pentru că variază în funcție de valorile altei sau altor variabile, care sunt denumite, din acest motiv, independente. Identificarea lor corectă în cazul unui studiu statistic este esențială pentru fundamentarea procedurilor statistice.

În esență, variabila dependentă face obiectul măsurării cu scopul de a fi supusă unor concluzii. Prin opoziție, variabila independentă este utilizată ca variabilă de influență, ale căror efecte posibile asupra variabilei

Copyright © DEPARTAMENT ID 2008

dependente urmează sa fie puse în evidență. Termenii „dependent", „independent" se utilizează în mod obişnuit în legătură cu cercetarea experimentală. În acest context există variabile „manipulate" adică „independente" de reacțiile, intențiile, conduitele sau trăirile subiecților investigați (toate acestea fiind variabile „dependente"). În raport cu analiza statistică, definirea variabilelor ca dependente şi independente nu este condiționată de măsurarea lor în condiții de experiment.

Nu există variabile care sunt „dependente" sau „independente" prin natura lor. Caracteristica de a fi de un tip sau de altul provine din rolul care le este atribuit de către cercetător într-un anumit context de cercetare. De exemplu, dacă presupunem că starea emoțională este influențată de fumat, rezultatul la un test de labilitate emoțională este variabila dependentă, iar fumatul, variabila independentă. Într-un alt studiu, însă, în care ne interesează frecvența fumatului în funcție de sex, numărul țigărilor este variabila dependentă, iar sexul, variabila independentă. Sexul, la rândul său, poate deveni variabilă dependentă într-un studiu privind relația dintre consumul unei anumite substanțe de către gravide şi sexului copiilor lor.

Sarcină de lucru nr. 1. 2

Identificați variabila independentă şi variabila independentă în următoarele situații:

Scrieți răspunsurile şi numai apoi verificați răspunsurile corecte la pagina 21

1. Timpul de studiu are un efect asupra rezultatelor şcolare.

v. dependentă _____ *v. independentă* _____

2. Medicația reduce simptomele depresiei.

v. dependentă _____ *v. independentă* _____

3. Zgomotul ambiant creşte nivelul de agresivitate.

v. dependentă _____ *v. independentă* _____

Variabile continue şi variabile discrete

Se numeşte „continuă" o variabilă de tip numeric care are un număr teoretic infinit de niveluri ale valorilor măsurate. Acest tip de variabilă poate lua, în principiu, orice valoare, permițând utilizarea zecimalelor. Exemple: timpul de reacție, înălțimea, greutatea

Se numeşte „discretă" o variabilă care prezintă un număr finit al valorilor pe care le poate lua (numărul persoanelor dintr-o familie, numărul de țigarete fumate zilnic).

Populație şi eşantion

Metoda ştiințifică permite studiul unui eşantion pentru a trage concluzii asupra populației din care este selecționat.

A fundamenta un adevăr statistic înseamnă a trage o concluzie care descrie parametrii unei populații de valori, pe baza indicatorilor unui eşantion din acea populație.

În contextul cercetării statistice utilizăm următoarele definiții:

Populație, totalitatea „unităților de informație" care constituie obiectivul de interes al unei investigații. Prin „unități individuale de informație" înțelegem cel mai adesea „persoane" (sau „subiecți", cu un termen uzual in cercetarea psihologică). Dar, la fel de bine, putem înțelege şi „populația de cupluri familiale", sau „populația" de diferențe dintre mediile a două variabile, de exemplu. În esență, prin „populație" trebuie să

Copyright © DEPARTAMENT ID 2008

înțelegem extinderea maximă posibilă, sub aspectul volumului, a respectivei „unități de informație". Extinderea menționată este, la rândul ei, definită prin obiectivul de cercetare, ceea ce înseamnă ca are o dimensiune subiectivă. Aceasta se referă la domeniul de interes pe care şi-l propune cercetătorul. De exemplu, într-un studiu cu privire la efectul oboselii asupra performanței cognitive, pot fi vizate diferite categorii de „populații": a aviatorilor, a studenților, a mecanicilor de locomotivă, a şahiştilor

Eşantion, reprezintă „unitățile de informație" selecționate pentru a fi efectiv studiate. Ideea pe care se bazează cercetările bazate pe eşantioane, este aceea că se pot face aprecieri asupra unei întregi populații, în anumite condiții, doar pe baza caracteristicilor măsurate pe o parte a acesteia.

Exemple:

- Într-un studiu asupra efectelor accesului la internet asupra elevilor de liceu, elevii de liceu reprezintă „populația", iar elevii selecționați pentru investigație, „eşantionul".
- Într-un studiu care vizează influența inteligenței asupra performanței în instruirea de zbor, populația este reprezentată de toți piloții, iar eşantionul, de subiecții incluşi în studiu.

Reprezentativitatea eşantionului este dată de calitatea valorilor acestuia de a descrie în mod corect caracteristicile populației din care a fost extras. Nici un eşantion nu poate reprezenta perfect datele populației. De aceea reprezentativitatea are o semnificație relativă. Ca urmare estimările pe bază de eşantion conțin întotdeauna o doză mai mare sau mai mică de eroare. Cu cât eroarea este mai mică, cu atât concluziile obținute pe eşantion pot fi generalizate mai sigur asupra populației. Pentru a permite fundamentarea inferențelor statistice, eşantionul trebuie să fie constituit din „unități de informație" (subiecți, valori) independente unele de altele.

Exemple:

- Dacă măsurăm timpul de reacție la un număr de cinci subiecți, dar facem trei evaluări la fiecare subiect, nu avem eşantion de 15 valori independente, deoarece valorile aceluiaşi subiect au în comun o „constantă personală" care le face dependente una de cealaltă. Pentru avea un singur eşantion am putea să utilizăm media celor trei determinări pentru fiecare subiect.
- Dacă dorim să investigăm efectul inteligenței asupra performanței şcolare, trebuie să avem grijă să includem în eşantion subiecți provenind din familii cu un nivel variat al veniturilor, pentru a anihila influența statutului socioeconomic asupra performanței şcolare.

Sarcină de lucru nr. 1. 3

Identificați eşantionul şi populația în următoarele situații:

Scrieți răspunsurile şi numai apoi verificați răspunsurile corecte

1. Un grup de studenți a fost selecționat dintre studenții de anul I.

*eşantion*_____ *populație* _____

2. La proiect au participat 100 de angajați ai companiei.

eşantion _____ *populație* _____

Copyright © DEPARTAMENT ID 2008

3. Sondajul a fost efectuat pe 1000 de persoane din România.
eşantion _____ *populație* _____

Statistica descriptivă şi statistica inferențială

Statistica descriptivă prezintă datele în formă numerică sau grafică.

Statistica descriptivă se referă la metodele cu ajutorul cărora analizăm caracteristicile variabilelor statistice. Dacă aplicăm un test de timp de reacție unui număr de 50 de persoane, putem calcula valoarea medie a timpilor de reacție, împrăştierea acestora sau, utilizând o tehnică de reprezentare grafică, modul în care se distribuie valorile prin raportare la un sistem de coordonate. Toate aceste prelucrări, şi altele încă, despre care vom vorbi pe larg mai departe, fac parte din categoria metodelor statisticii descriptive.

Statistica inferențială aplică proceduri de decizie cu privire la adevărul unei ipoteze.

Statistica inferențială cuprinde metodele de verificare a ipotezelor de cercetare prin testarea ipotezelor statistice. Să presupunem că cei 50 de subiecți de mai sus sunt supuşi aceluiaşi test de tip de reacție în condiții de noxe de mediu (de exemplu, zgomot excesiv) pentru a verifica ipoteza că zgomotul reduce promptitudinea reacțiilor.

Statistica parametrică şi statistica neparametrică

Procedurile parametrice testează variabile cantitative. Procedurile neparametrice testează variabile calitative.

Esența procedurilor statistice este verificarea ipotezelor. Aceasta se face prin utilizarea unor proceduri de calcul care urmăresc punerea în evidență a legăturilor dintre variabile. Atunci când aceste proceduri se aplică unor situații în care variabilele dependente sunt de tip cantitativ (interval/raport), procedura se numeşte „parametrică”. Prin opoziție, procedurile aplicate în cazul în care variabilele dependente sunt de tip „calitativ” (nominale sau ordinale) se numesc „neparametrice”.

Studii experimentale şi studii observaționale

Studiile de tip corelațional evidențiază relații între variabile, dar nu permit concluzii de tip cauzal.

În cazul **studiilor experimentale**, cercetătorul nu se limitează la măsurarea variabilei independente ci o şi manipulează. De exemplu, dacă analizăm rezultatele a două grupe de trăgători la țintă, unii care au efectuat în prealabil şedințe de relaxare şi alții care nu au efectuat, avem de a face cu un studiu numit „corelațional”. Pe baza lui putem constata dacă există o legătură între cele două variabile, dar în nici un caz dacă relaxarea determină („cauzează”) creşterea performanțelor.

Studiile de tip experimental evidențiază relații de tip cauzal între variabile.

În cazul studiilor numite observaționale, variabilele dependente şi independente sunt măsurate în condiții care nu permit concluzii de tip cauzal. Aplicarea unui test de personalitate unor categorii de subiecți, diferite în funcție de sex sau vârstă, de exemplu, urmată de compararea rezultatelor între categorii şi constatarea existenței unor diferențe, fie şi semnificative statistic, nu înseamnă că personalitatea este „influențată” de

Copyright © DEPARTAMENT ID 2008

apartenența la o anumită categorie. Totuşi, rezultatele studiilor „corelaționale" pot fi interpretate uneori în termeni cauzali, utilizând teorii existente sau ipoteze, dar astfel de rezultate nu pot constitui în nici un caz o dovadă a unei relații de tip cauzal.

Rezumatul unității de învățare

- Statistica este disciplina care se ocupă cu sintetizarea, prezentarea şi analiza datelor numerice, în scopul evidențierii semnificațiilor acestora.
- Statistica este un instrument al metodei ştiințifice în psihologie.
- Componentele metodei ştiințifice sunt: observația – elaborarea ipotezei – analiza datelor empirice – concluzia
- Măsurarea înseamnă a atribui numere sau simboluri unor caracteristici ale realității obiective sau subiective, în funcție de anumite aspecte cantitative sau calitative care le caracterizează.
- Măsurarea pe scală nominală, identifică prezența unei anumite caracteristici, fără a avea o semnificație cantitativă. Variabilele nominale se referă la caracteristici calitative şi categoriale.
- Măsurarea pe scală ordinală, identifică raportul de ordine între valori, fără a preciza distanța cantitativă dintre acestea. Variabilele ordinale se referă la caracteristici calitative.
- Măsurarea pe scală de interval, aduce în plus față de scala ordinală precizarea distanței dintre ranguri. Din acest motiv este o scală de tip cantitativ.
- Măsurarea pe scală de raport aduce în plus față de scala de interval, raportarea la o valoare minimă absolută.
- Statistica descriptivă se ocupă cu sintetizarea şi prezentarea datelor în timp ce statistica inferențială se ocupă cu generalizarea rezultatelor dincolo de eşantionul pe care au fost recoltate.
- Variabilele dependente sunt cele care fac obiectul interesului direct al cercetătorului, fiind măsurate în vederea extragerii unei concluzii. Variabilele independente reprezintă condiția sau contextul din care rezultă variația valorilor variabilei dependente.
- Atunci când variabila dependentă implicată într-un studiu statistic este măsurată pe o scală de tip calitativ (nominal sau ordinal), se aplică una dintre procedurile statistice neparametrice. În cazul variabilelor măsurate pe scale cantitative se aplică, de regulă, statistici parametrice, fără ca acest lucru să fie posibil întotdeauna.
- Studiile de tip corelațional pun în evidență relația dintre variabile fără a susține concluzii de tip cauzal. Studiile de tip experimental pun în evidență relații de tip cauzal între variabile.

Copyright © DEPARTAMENT ID 2008

Răspunsuri corecte la sarcinile de lucru

Tema de reflecție nr. 1:
Răspuns: scală ordinală
***Comentarii*.** Nivelul de măsurare creşte progresiv de la scala de tip nominal la cea de tip raport, dar cuantificarea exactă a acestei este imposibilă.

Sarcina de lucru nr. 1.1

numărul întrebării	Răspuns
1	nominală
2	raport
3	ordinală
4	raport
5	ordinală

Comentarii. Scala pe care este evaluată o variabilă se defineşte în funcție de modul de atribuire a valorilor. Astfel, este posibil ca, în funcție de acest lucru, o anumită variabilă să fie exprimată pe scale diferite.

Sarcina de lucru nr. 1.2

4. Timpul de studiu are un efect asupra rezultatelor şcolare.
   v.dependentă: rezultatele şcolare
   v. independentă: timpul de studiu
5. Medicația reduce simptomele depresiei.
   v.dependentă: simptomele depresiei
   v. independentă: medicația
6. Zgomotul ambiant creşte nivelul de agresivitate.
   v.dependentă: nivelul de agresivitate
   v. independentă: zgomotul ambiant

Comentarii. În studiile de tip corelațional, identificarea variabilei dependente şi a variabilei independente se va face prin plasarea lor mintală într-o relație de tip cauzal, fără ca rezultatele studiului să poată fi interpretate în mod cauzal.

Sarcina de lucru nr. 1.3
4. Un grup de studenți a fost selecționat dintre studenții de anul I.
   eşantion: grupul de studenți
   populație: studenții anului I
5. La proiect au participat 100 de angajați ai companiei.
   eşantion 100 de angajați:
   populație: toți angajații companiei
6. Sondajul a fost efectuat pe 1000 de persoane din România.
   eşantion: 1000 de persoane
   populație: toată populația României

Copyright © DEPARTAMENT ID 2008

Comentarii. Se va observa că, de fiecare dată, populația studiului este diferită ca mărime, în funcție de nivelul de generalizare pe care cercetătorul doreşte să îl dea rezultatelor.

Lucrarea de evaluare nr. 1.1

Lucrarea de evaluare va fi publicată pe portal (http://portal.credis.ro). Data limită de trimitere este preziua tutorialului. După acest termen lucrările nu mai sunt acceptate.

Bibliografie

Bibliografia de bază

- Marian Popa, (2008), Statistică pentru psihologie. Teorie şi aplicații SPSS, editura Polirom
- Pagina web a cursului, la adresa www.mpopa.ro

Bibliografie suplimentară

- Clocotici, V., & Stan, A. (2000). Statistica aplicata in psihologie. Iasi: Polirom, p. 11-57
- Rotaru, T. (coord.). (1999). Metode statistice aplicate in stiintele sociale. Iasi: Polirom. p. 15-28
- Radu I., (coord), (1993), Metodologie psihologică şi analiza datelor, Editura Sincron, p. 45-51
- Vasilescu, I. P. (1992). Statistica informatizata pentru stiinte despre om (Vol. 1-2). Bucuresti: Editura militara., p.5-43

Copyright © DEPARTAMENT ID 2008

STATISTICI DESCRIPTIVE

Obiective de învățare şi informații introductive

Parcurgerea acestei unități, va permite studenților:

- ✓ să utilizeze tehnicile numerice de analiză globală a variabilelor statistice (analiza de frecvențe);
- ✓ să utilizeze tehnicile grafice de analiză a variabilelor statistice (histograma, graficul de tip bară, graficul circular, reprezentarea *stem-and-leaf*);
- ✓ să calculeze indicatorii tendinței centrale (modul, mediana, media)
- ✓ să calculeze indicatorii împrăştierii (amplitudinea, abaterea quartilă, abaterea medie, abaterea standard, coeficientul de variație);
- ✓ să utilizeze indicatorii formei distribuției (simetrie şi boltire):
- ✓ să analizeze valorile extreme ale distribuțiilor statistice.

Definirea şi componentele statisticii descriptive

istica descriptivă are drept obiective organizarea, sintetizarea şi descrierea datelor. Rezultatul măsurării se traduce în obținerea unei colecții de date. Să presupunem că am aplicat un test de cunoştințe unui grup de 25 de studenți şi am obținut următoarea distribuție de valori pentru variabila „răspunsuri corecte”:

8, 6, 10, 9, 6, 6, 8, 7, 4, 9, 6, 2, 8, 6, 10, 4, 5, 6, 8, 4, 7, 8, 4, 7, 6

Datele de mai sus reprezintă valorile variabilei statistice „răspunsuri corecte” (denumite şi „serie statistică” sau „distribuție statistică”), care este compusă din 25 de „valori” sau „scoruri”. Fiind rezultatul primar al măsurării, aceste valori se mai numesc şi valori „primare” sau „brute”. Valorile acestei variabile sunt exprimate pe o scală cantitativă de tip raport.

Privite sub forma în care se prezintă mai sus, datele respective ne spun puține lucruri. Iar dacă ar fi şi mai multe, de ordinul sutelor sau miilor, atunci ar fi practic imposibil de făcut vreo apreciere, în această formă de prezentare . De aceea, pentru a ne face o imagine mai coerentă asupra unei serii de valori, acestea trebuie supuse unor operații care să scoată în evidență caracteristicile distribuției

Definiție: *Tehnicile şi procedurile destinate organizării şi prezentării sumative a datelor, constituie ceea ce se numeşte statistica descriptivă.*

Principalele componente ale statisticii descriptive sunt:

- Tehnici de organizare şi prezentare a datelor, care pot fi, la rândul lor:
  - numerice (distribuția de frecvențe simple sau grupate;)
  - grafice (histograme; grafice de tip bară, linie, circular, histograma *stem-and-leaf*)
- Indicatori numerici sumativi, care sunt la rândul lor de trei tipuri:
  - indicatori ai tendinței centrale (mod, medie, mediană)
  - indicatori ai împrăştierii (amplitudine, abatere quartilă, abatere standard)
  - indicatori ai formei distribuției (simetrie şi boltire).

Copyright © DEPARTAMENT ID 2008

Dincolo de scopul în sine al acestor proceduri, acela de a oferi o imagine sintetică asupra datelor analizate, trebuie să înțelegem statistica descriptivă şi ca pe o etapă pregătitoare în fundamentarea procedurilor statisticii inferențiale (destinată verificării ipotezelor statistice) despre care vom vorbi mai târziu.

Statistici descriptive globale

Tehnicile descriptive de tip global se referă la prezentarea şi analiza tuturor valorilor unei distribuții statistice. Aceste tehnici sunt, la rândul lor de două feluri: numerice (analiza de frecvențe) şi grafice.

Analiza de frecvențe

Analiza de frecvențe simple

Analiza de frecvențe simple se bazează pe frecvența de apariție a fiecărei valori dintr-o distribuție

Dacă ne întoarcem la seria de valori de mai sus, cel mai simplu lucru pe care putem să îl facem, şi care ne poate da o anumită imagine asupra ei, este sortarea, punerea valorilor în ordine crescătoare sau descrescătoare:

10, 10, 9, 9, 8, 8, 8, 8, 8, 7, 7, 7, 6, 6, 6, 6, 6, 6, 6, 5, 4, 4, 4, 4, 2

Privind datele aranjate ca mai sus putem observa cu uşurință câteva lucruri: valoarea cea mai mare şi valoarea cea mai mică, valorile care se repetă. Dar, chiar şi acest mod de prezentare, nu ne-ar fi de mare ajutor dacă valorile ar fi într-un număr mare. Pentru a elimina acest neajuns se foloseşte tabelul frecvențelor simple.

Tabelul 1. Frecvențe simple

Valoare	fa
10	2
9	2
8	5
7	3
6	7
5	1
4	4
3	0
2	1
Total	**Σfa=25**

Dacă luăm în considerare seria de valori de mai sus, un tabel al frecvențelor simple (absolute) este compus din lista valorilor distincte, ordonate descrescător, la care se adaugă frecvența absolută (**fa**) a fiecărei valori (de câte ori se întâlneşte în cadrul seriei).

Copyright © DEPARTAMENT ID 2008

Se observă că astfel datele au un caracter mai ordonat, iar coloana frecvențelor absolute scoate în evidență anumite aspecte cum ar fi, de exemplu, faptul că cea mai frecventă valoare este 6 (apare de 7 ori). Observăm că seria de valori din tabel include toate valorile posibile între valoarea cea mai mare (10) şi cea mai mică (2), incluzând şi valorile care nu se întâlnesc în mod real în cadrul seriei. În cazul nostru avem valoarea 3, cu frecvența de apariție 0. Suma frecvențelor absolute (Σfa) indică totalul valorilor din cadrul seriei (25).

În practică, pe lângă frecvențele absolute se iau în considerare şi alte tipuri de frecvențe (vezi tabelul 2):

- **Frecvența cumulată (fc).** Totalul valorilor care se cumulează începând de la valoarea cea mai mare până la valoarea cea mai mică din tabel. De exemplu, în tabelul sintetic de mai jos, avem 6 valori mai mici sau egale cu 5, 21 de valori mai mici sau egale cu 8 şi, evident, 25 de valori mai mici sau egale cu 10.
- **Frecvența relativă raportată la unitate fr(1).** Este raportul dintre frecvența absolută şi suma frecvențelor absolute (fa/Σfa).

Exemple:

- pentru valoarea 10: fa/Σfa=2/25=0.08;
- pentru valoarea 6: fa/Σfa=7/25=0.13; ş.a.m.d.

- **Frecvența relativă cumulată, raportată la unitate fr(1)**: Este similară frecvenței cumulate absolute, cu deosebirea că în acest caz se cumulează frecvențele relative.

Exemple:

- Dacă privim întreaga serie ca întreg (egală cu 1 sau „unitate"), atunci toate valorile mai mici sau egale cu 5 au o frecvență cumulată egală cu 0.24 (adică, fr(1)=0.04+0+0.16+0.04=0.24)
- Pentru valoarea 7, frecvența relativă cumulată raportată la unitate este: frc(1)=0.04+0+0.16+0.04+0.28+0.12=0.64
- Frecvența relativă cumulată pentru valoarea cea mai mare din serie este întotdeauna 1.00 (corespunzătoare în cazul nostru valorii 10).

- **Frecvența relativă procentuală fr(%):** Exprimă procentul valorilor care se situează până la o anumită valoare din cadrul distribuției. Se calculează fie prin înmulțirea fr(1) cu 100, fie prin calcularea directă procentului pe care îl reprezintă o anumită valoare raportat la totalul valorilor dintr-o distribuție. Suma frecvențelor relative procentuale este întotdeauna egală cu 100.

Exemple (tabelul 2):

- 8% dintre studenții evaluați au realizat 10 răspunsuri corecte
- 28% dintre studenții evaluați au realizat 6 răspunsuri corecte

- **Frecvența relativă cumulată procentuală (frc%)**: Exprimă procentul valorilor dintr-o distribuție care se plasează până la o anumită valoare (inclusiv aceasta).

Exemple:

- 52% dintre studenți au obținut o notă egală sau mai mică de 6
- 92% au obținut cel puțin nota 9
- Desigur, pentru valoarea maximă a unei distribuții, frecvența cumulată procentuală este întotdeauna 100%.
  - Frecvența relativă procentuală cumulată se numeşte **rang percentil**. Astfel, despre valoarea 6 din distribuția de mai sus se poate

Copyright © DEPARTAMENT ID 2008

spune că are rangul percentil 52, adică, 52% dintre valorile unei distribuții sunt între cea mai mică valoare şi valoarea 6, inclusiv.

- Prin convenție, rangul percentil se defineşte ca procentajul datelor valorilor dintr-o distribuție care se află până la o anumită valoare inclusiv.
- În mod complementar, numim **percentilă**, valoarea dintr-o distribuție care corespunde unui anumit rang percentil. În exemplul de mai sus, rangului percentil 52 îi corespunde valoarea 6, numită, de aceea, percentila 52.
- În practică, există anumite percentile care au o importanță aparte. Acestea sunt percentilele corespunzătoare rangurilor percentile cu valorile 10, 20, 30,..., 100. Despre semnificația lor vom vorbi mai târziu în acest curs. De asemenea, se utilizează termenul de quartile pentru percentilele care împart distribuția în patru zone egale ca număr de valori. Acestea sunt corespunzătoare rangurilor percentile de 25, 50 şi 75. Cu alte cuvinte, valoarea dintr-o distribuție până la care se află 25% din valori este percentila 25, valoarea până la care se află 50% este percentila 50, iar valoarea până la care se află 75% din valori este percentila 75.

Tabelul 2. Tabloul sintetic al frecvențelor simple

Valoare	fa	fc	fr (1)	frc (1)	fr (%)	frc (%)
10	2	25	0,08	1,00	8%	100%
9	2	23	0,08	0,92	8%	92%
8	5	21	0,20	0,84	20%	84%
7	3	16	0,12	0,64	12%	64%
6	7	13	0,28	0,52	28%	52%
5	1	6	0,04	0,24	4%	24%
4	4	5	0,16	0,20	16%	20%
3	0	1	0	0,04	0%	4%
2	1	1	0,04	0,04	4%	4%
Total	**Σfa=25**		**Σfr=1**		**Σfr%=100**	

Analiza de frecvențe grupate

Analiza de frecvențe grupate se bazează pe frecvența de apariție a claselor (grupelor) de valori într-o distribuție

Aranjarea unei distribuții sub forma tabelului de frecvențe simple este foarte utilă dar nu este practică atunci când avem o distribuție, cu un număr mare sau foarte mare de valori, care ar genera un tabel cu prea multe linii pentru a fi inteligibil.

Să presupunem că valorile de mai jos reprezintă distribuția variabilei „inteligență" măsurată prin aplicarea unui test la un număr de 50 de subiecți.

101	94	87	117	115	116	91	113	96	105
92	107	118	114	98	112	101	114	107	109
97	109	124	102	118	113	116	106	108	89
106	108	115	92	97	102	108	102	109	114
107	104	110	101	101	121	125	86	109	123

Datele din tabel sunt aranjate la întâmplare, analiza lor fiind dificilă. Presupunând că le-am ordona şi am face tabelul frecvențelor simple, am

Copyright © DEPARTAMENT ID 2008

obține un uşor progres, dar încă ar fi greu de analizat deoarece vom obține un tabel cu prea multe valori distincte.

Pentru a ne face o imagine sintetică a distribuției, ne propunem să realizăm un număr de categorii (clase) cuprinse între anumite intervale de performanță la test, urmând să stabilim apoi care este frecvența de apariție a fiecărei clase în distribuția noastră. Această tehnică de organizare a datelor se numeşte „frecvența grupată”

Pentru a realiza un tabel de frecvențe grupate se procedează astfel:

1. Alegem numărul de intervale (clase, categorii), recomandabil, între 5 şi 15 (valori stabilite convențional şi orientativ)
2. Definim mărimea intervalului de clasă, respectând următoarele reguli:
   - toate intervalele trebuie să fie egale
   - limitele intervalelor trebuie să cuprindă toate valorile (între limitele intervalelor alăturate să nu existe „goluri” sau suprapuneri)

Pentru distribuția de mai sus, paşii de realizare a analizei de frecvențe grupate se concretizează astfel:

Se face diferența dintre valoarea cea mai mare şi valoarea cea mai mică	125 – 86 = 39
Se împarte valoarea obținută la mărimea posibilă a intervalului de clasă (2, 3, 5 sau 10) pentru a realiza numărul de clase al noii distribuții	39/2 = ~20 clase (prea multe) 39/3 = 13 clase (variantă posibilă) 39/5 = ~ 8 clase (variantă acceptabilă)
Se selectează mărimea intervalului care conduce la un număr de clase cuprins între 5 şi 15.	Vom alege 5, pentru că produce o distribuție cu 8 clase care este mai uşor de analizat şi manipulat
Se determină limita inferioară a primului interval (trebuie să fie un multiplu al mărimii intervalului)	Alegem valoarea 85 ca limită inferioară
Se determină limita superioară a primului interval	Dacă mărimea intervalului este 5, limita superioară va fi 89 (85,86,87,88,89)
Se construiesc intervalele de clasă pentru fiecare interval (vezi coloana „clase” din tabelul 3)	
Se aplică analiza de frecvențe ca în cazul frecvențelor simple, aplicată la clase	

În fine, alegerea dimensiunii intervalului trebuie să țină seama şi de caracteristicile distribuției simple (discutată anterior). Intervalele trebuie astfel alese încât să se evite situația de a avea clase care cuprind un număr excesiv de valori în timp ce altele sunt puțin reprezentate sau nu conțin nici o valoare.

În exemplul dat, deşi valoarea maximă a variabilei este 125, intervalul maxim este 125-129, deoarece intervalele declarate trebuie să fie egale. Ca urmare, tabelul frecvențelor grupate va arăta astfel:

Tabelul 3. Tabelul de frecvențe grupate

Clase	fa	fr%	frc%
125 – 129	1	2%	100%
120 – 124	3	6%	98%

Copyright © DEPARTAMENT ID 2008

115 – 119	7	14%	92%
110 – 114	7	14%	78%
105 – 109	13	26%	64%
100 – 104	8	16%	38%
95 – 99	4	8%	22%
90 – 94	4	8%	14%
85 – 89	3	6%	6%
	Σfa=50	**Σfr%=100**	

Este de la sine înțeles că clasele de intervale (grupele) vor putea fi analizate într-o manieră similară frecvențelor simple, utilizând valorile absolute (fa) sau valorile relative raportate la unitate sau procentuale (fr(1), fr%). Analizând tabelul de mai sus, putem observa că cei mai mulți subiecți au obținut un scor la testul de inteligență cuprins între 105 şi 109 (fa=13), aceştia reprezentând 26% din totalul subiecților evaluați. În fine, din coloana frecvențelor relative procentuale cumulate putem deduce că 64% dintre subiecți obțin o performanță de maxim 109 sau mai mică (sau, dacă dorim, 36 % dintre subiecți obțin o performanță de minim 105) etc.

Sarcina de lucru nr. 2.1

Alegeți varianta de răspuns aleasă sau scrieți răspunsul în text, apoi verificați răspunsurile corecte

1. Percentila 25 este acea valoare a unei distribuții care:
   a. are 75% din valori mai mari decât ea
   b. se întâlneşte la 25% dintre subiecți
   c. împarte distribuția în 25 de părți egale
   d. nici una din variantele de mai sus
2. Percentila 50 este o valoare identică cu:
   a. quartila 3; b. quartila 1; c. mediana; d. abaterea standard
3. Ce procent de valori este reprezentat în caseta reprezentării *box-plot*:
   a. 50%; b. 25%; c. 30%; d. 75%
4. Ce reprezintă frecvența relativă raportată la unitate?

   _____
5. Ce înseamnă faptul că pe coloana frecvenței relative procentuale din dreptul unui anumite valori este scris 7%?

   _____
6. Cum se stabileşte limita inferioară a primei clase, în cazul unei distribuții de frecvențe grupate?

   _____
7. Care este numărul recomandabil de clase într-o distribuție de frecvențe grupate?

   _____
8. Cum se numesc valorile de pe coloana frecvențelor relative procentuale cumulate?

   _____

Copyright © DEPARTAMENT ID 2008

9. Cum se numeşte valoarea variabilei care corespunde unui anumit rang percentil?

Reprezentarea grafică a datelor

Graficele sunt imagini ale distribuțiilor de frecvențe.

Reprezentările graficele sunt forme intuitive de prezentare a distribuțiilor de frecvențe („o imagine *face mai mult decât o mie de cuvinte*"). Ele sunt foarte frecvent utilizate pentru analiza şi prezentarea datelor în psihologia aplicată deoarece facilitează înțelegerea semnificației datelor numerice. În prezent, programele computerizate oferă mijloace extrem de puternice şi de sofisticate pentru elaborarea reprezentărilor grafice. Dar simpla utilizare a unui astfel de program nu garantează realizarea unui grafic eficient. În esență, un grafic eficient este o combinație reuşită între formă şi conținutul statistic pe care îl reflectă. Realizarea acestei combinații depinde de respectarea câtorva principii esențiale:

- ✓ focalizarea pe conținutul şi nu pe forma graficului
- ✓ este esențial să fie evitate distorsiunile induse de forma graficului
- ✓ este recomandabil să fie utilizate grafice care favorizează comparații între variabile şi nu doar reprezentări individuale, "statice", ale acestora
- ✓ fiecare grafic trebuie să servească un singur scop, exprimat clar şi evident
- ✓ orice grafic va fi însoțit de informații statistice şi descrierile necesare pentru a fi uşor şi corect înțeles
- ✓ un grafic trebuie să scoată în evidență datele şi nu abilitățile tehnice de editare ale celui care l-a creat.

Formele de expresie grafică a datelor statistice sunt foarte numeroase. Ne vom ocupa aici doar de câteva dintre acestea, cel mai des utilizate[1]:

- graficul de tip bară
- histograma
- poligonul de frecvențe
- graficul frecvenței cumulate
- graficul circular
- graficul de tip „*stem and leaf*" („tulpină şi frunze")

<u>Graficul de tip bară</u>

Este cel mai simplu mod de reprezentare grafică a datelor. Se utilizează atunci când dorim să reprezentăm o variabilă „discretă" (care prezintă valori întregi, de exemplu, numărul de răspunsuri corecte la un test în funcție de nivelul de instruire al subiecților).

În mod obişnuit, un grafic se prezintă ca o imagine inclusă într-un sistem de axe perpendiculare:

- Axa orizontală (Ox) pe care sunt reprezentate valorile distribuției

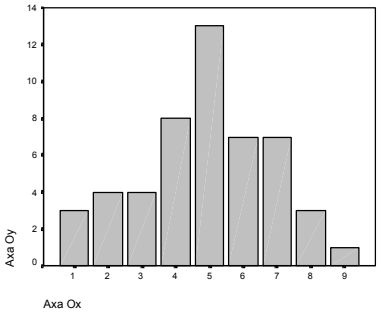
[1] O prezentare extensivă a tipurilor de reprezentări grafice poate fi găsită în Statistica, Electronic Textbook, 1984-1999, ©StatSoft Inc., Graphical techniques

Copyright © DEPARTAMENT ID 2008

- Axa verticală (Oy) pe care sunt reprezentate frecvențele

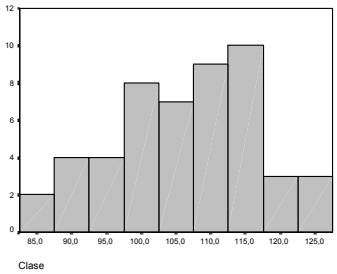
fiecărei valori, sub forma unei bare rectangulare.

Iată cum arată un grafic de acest tip efectuat pe datele din tabelul de frecvențe grupate, luând clasele drept valori ale distribuției. Cu cât frecvența unei valori este mai mare, cu atât bara este mai mare. Simplitatea şi claritatea este cea mai mare calitate a acestui tip de grafic.



Histograma

La prima vedere, histograma este asemănătoare cu graficul de tip bară. Ea este mai adecvată pentru situațiile când variabila pe care dorim să o reprezentăm este de tip „continuu” (adică poate lua orice valoare pe o scală numerică, de ex., număr de răspunsuri corecte, timpul de reacție, lungimea). Iată, de exemplu, histograma distribuției de frecvențe din tabelul 3 (realizată cu programul SPSS):



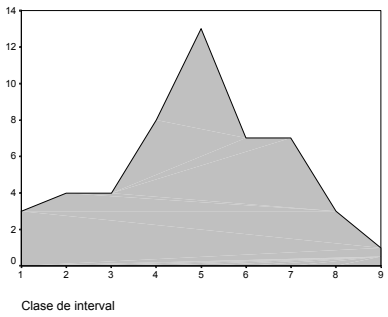
Se observă faptul că programul a realizat automat o grupare de frecvențe, afişând pe axa Ox limita minimă a intervalului ca „etichetă” a acestuia.

În principiu, nimic nu ne împiedică să realizăm o histogramă pe aceleaşi valori care au fost reprezentate pe un grafic de tip bară.

Poligonul de frecvențe

Este o reprezentare alternativă la histogramă. Punctele centrale ale suprafețelor rectangulare care reprezintă frecvența sunt unite cu o linie care delimitează suprafața poligonului.

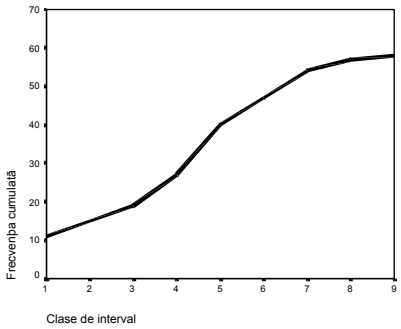
Copyright © DEPARTAMENT ID 2008



Poligonul alăturat prezintă distribuția de frecvențe grupate din tabelul de mai sus, cifrele 1,2,3,4,5,6,7,8,9 reprezentând denumirea convențională a fiecărei clase.

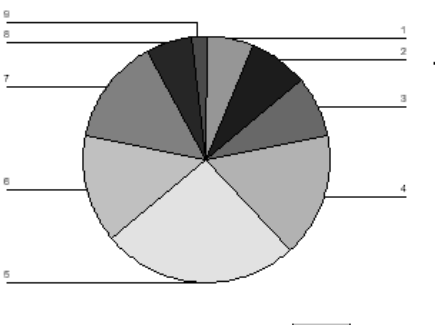
Graficul frecvenței cumulate

Este un grafic de tip liniar care reprezintă valorile frecvenței absolute cumulate. Pe acest grafic se vede cu uşurință câte valori se află până la o anumită valoare din distribuție (datele reprezentate sunt cele din tabelul 3, fiecare interval de clasa fiind etichetat convențional cu cifre de la 1 la 9).



Graficul circular

Este utilizat în situațiile în care valorile sunt „parte a unui întreg”. De exemplu, poate fi utilizat la reprezentarea distribuției de frecvențe grupate de mai sus, pentru a avea o imagine directă a ponderii frecvenței fiecărei clase de interval în raport cu celelalte.



Graficul alăturat reprezintă frecvența absolută a claselor de interval ale aceleiaşi distribuții de mai sus. Pe un grafic de acest tip se pot

Copyright © DEPARTAMENT ID 2008

reprezenta fie valorile absolute, fie procentajul fiecărei clase raportat la întreg.

Reprezentarea de tip *stem-and-leaf* (*stem plot*)

Este o reprezentare care încearcă să îmbine expresia numerică cu cea grafică, fiind propusă de statisticianul J.W. Tuckey (1977). Scopul principal a fost acela de a oferi nu doar o imagine a distribuției ci şi o metodă de explorare a acesteia. Ea este din ce în ce mai utilizată de psihologi, motiv pentru care considerăm necesar să o prezentăm aici.

Atunci când utilizăm o distribuție de frecvențe grupate, cazurile individuale „se pierd" la nivelul fiecărei clase de interval fără a mai putea şti unde se plasează fiecare valoare inițială în interiorul fiecărui interval. Reprezentarea de tip *stem-and-leaf* (pe scurt *stem plot*), are tocmai avantajul de a realiza graficul distribuției cu păstrarea valorilor individuale.

Modul de realizare

Să revenim la distribuția prezentată anterior:

101	94	87	117	115	116	91	113	96	105
92	107	118	114	98	112	101	114	107	109
97	109	124	102	118	113	116	106	108	89
106	108	115	92	97	102	108	102	109	114
107	104	110	101	101	121	125	86	109	123

Mai întâi, observăm că valorile sunt cuprinse între 86 şi 125. Alegem o valoare convenabilă pentru tulpină, care va juca rolul de interval de clasa, care în cazul nostru poate fi 10. „Tulpina" reprezentării *stem plot* este în acest caz numărul de zeci din fiecare valoare individuală.

Stem-and-Leaf

```
   8 .  679
   9 .  1224
   9 .  6778
  10 .  11112224
  10 .  5667778889999
  11 .  0233444
  11 .  5566788
  12 .  134
  12 .  5
 Mărimea „tulpinii":   10
```

Valorile din coloana *stem* indică numărul de zeci, iar cele din coloana *Leaf*, numărul de unități. Dacă privim imaginea în ansamblu ne-o putem reprezenta ca pe o histogramă orizontală. În acest exemplu:

Stem 8, urmat de *Leaf* 679 indică faptul că variabila noastră are în compunere valorile 86,87,89.

Stem 12, urmat de *leaf* 134, ne arată că distribuția conține valorile 121, 123,124

Copyright © DEPARTAMENT ID 2008

Sarcina de lucru nr. 2.2

Scrieți răspunsul în text, apoi verificați răspunsurile corecte

1. Pentru ce scale de măsurare se utilizează graficul de tip histogramă?

_____

2. Prin ce se deosebeşte graficul de tip *stem-and-leaf* de histograma?

_____

3. În ce situație se utilizează graficul de tip circular?

_____

4. Faceți reprezentarea *stem-and-leaf* pentru următoarea distribuție de valori:
   29, 28, 36, 41, 25, 15, 33, 40, 33, 20, 35, 26, 32, 23

Indicatori statistici descriptivi

Tipuri de indicatori sintetici:

Un indicator statistic concentrează într-o singură valoare o anumită caracteristică a distribuției

Trei sunt caracteristicile distribuțiilor care sunt evaluate cu ajutorul indicatorilor sintetici: tendința centrală, variabilitatea (împrăştierea, diversitatea), forma distribuției. Pentru fiecare din aceste caracteristici se utilizează anumiți indicatori specifici:

- Indicatori ai tendinței centrale: Aceştia sunt valori tipice, reprezentative, care descriu distribuția în întregul ei;
- Indicatori ai variabilității: Sunt valori care descriu caracteristica de împrăştiere a distribuției. O distribuție care conține aceeaşi valoare, ori de câte ori s-ar repeta ea, are o variabilitate zero.
- Indicatori ai formei distribuției: Sunt valori care se referă la forma curbei de reprezentare grafică a distribuției, prin comparație cu o curbă normală (oblicitate, aplatizare)

Indicatori ai tendinței centrale

Modul (Mo)

Este expresia ce mai directă a valorii tipice (reprezentative)a unei distribuții statistice.

În cazul unei distribuții simple, este valoarea cu frecvența cea mai mare de apariție

Copyright © DEPARTAMENT ID 2008

În cazul unei distribuții de frecvențe grupate, este clasa de interval cu frecvența cea mai mare de apariție

Modul se află prin alcătuirea tabelei de frecvențe (simple sau grupate) şi este identificat ca valoarea căreia îi corespunde frecvența absolută cea mai ridicată.

Distribuțiile pot avea un singur mod (*unimodale*), două moduri (*bimodale*) sau mai multe (*multimodale*)

Exemplu: În seria de valori 5,8,3,2,5,4, Mo=5 (apare de cele mai multe ori)

Mediana (Me)

Este valoarea „din mijlocul" unei distribuții, adică aceea care are 50% dintre valori deasupra ei şi 50% dintre valori dedesubtul ei (cu alte cuvinte, percentila 50).

Se găseşte prin alcătuirea tabelei de frecvențe, în coloana frecvențelor relative procentuale cumulate, şi corespunde valorii de 50%.

În cazul distribuțiilor cu număr impar de valori, Me este chiar valoarea respectivă.

În cazul distribuțiilor pare, Me se calculează ca medie a celor două valori din mijlocul distribuției

Exemplu: În seria de valori 5,8,3,2,5,4, ordonată crescător (2,3,4,5,5,8), Me=4,5 (ca medie a valorilor 4 şi 5 aflate în mijlocul unei distribuții pare). Dacă distribuția noastră ar fi avut 5 valori (fără 2, de exemplu), Me=5

Media aritmetică (m)

Media este cel mai utilizat indicator al tendinței centrale

Este raportul dintre suma valorilor distribuției şi numărul acestora

Notații uzuale:

- μ (*miu*), atunci când este media întregii populații de referință
- *m*, atunci când se calculează pentru un eşantion (cazul cel mai frecvent)

Calcularea mediei pentru o distribuție simplă de frecvențe se face prin adunarea valorilor şi se împărțirea la numărul lor

Exemplu: Pentru distribuția 5,8,3,2,5,4

$$m = \frac{\sum X}{N} = \frac{5+8+3+2+5+4}{6} = \frac{26}{6} = 4{,}50 \qquad \text{(formula 2.1)}$$

Calcularea mediei pentru o distribuție de frecvențe grupate: Se face suma produsului dintre fiecare valoare şi frecvența ei, apoi se împarte la suma frecvențelor (numărul valorilor)

Exemplu: Pentru distribuția: 5,8,3,3,3,2,4,2,3,5,4

$$m = \frac{\sum (X * f)}{\sum f} = \frac{5*2+8*1+3*4+2*2+4*2}{2+1+4+2+2} = \frac{43}{11} = 3{,}90$$

(formula 2.2)

NOTĂ: În expresia de mai sus:

- X este variabila.

Copyright © DEPARTAMENT ID 2008

- Prin $\sum X$ se înțelege ca „*Sumă de la X=1 la N* (numărul valorilor)
- f este frecvența . $\sum f$ se înțelege ca „*Sumă de la f=1 la k* (unde k numărul grupelor de frecvență)

Proprietățile mediei aritmetice

- Adăugarea\scăderea unei constante la fiecare valoare a distribuției, măreşte\scade media cu acea valoare
- Înmulțirea\împărțirea fiecărei valori a distribuției cu o constantă, multiplică\divide media cu acea constantă
- Suma abaterii valorilor de la medie este întotdeauna egală cu zero
- Suma pătratului abaterilor de la medie va fi întotdeauna mai mică decât suma pătratelor abaterilor în raport cu oricare alt punct al distribuției

Valori nedeterminate şi clase deschise

Valorile „nedeterminate" sunt acele valori a căror mărime nu decurge din procesul de măsurare, în acelaşi mod în care rezultă oricare valoare a seriei (Exemplu: La testul de asociere verbală, dacă subiectul depăşeşte, să zicem 10 sec., se înregistrează valoarea 10, fără a se aştepta, la infinit (?), un răspuns). Categorii „deschise" sunt acele categorii de valori care au una dintre limite „liberă" (Exemplu: *Câte țigări fumezi zilnic*? Se poate înregistra numărul țigărilor ca atare, dar ultima valoare este „30 sau mai mult).

În ambele situații de mai sus, utilizarea mediei este nesigură (şi incorectă). Indicatorul recomandabil este mediana.

Avantajele şi dezavantajele indicatorilor tendinței centrale

Tabloul de mai jos prezintă, în mod sintetic avantajele şi dezavantajele specifice indicatorilor tendinței centrale:

	AVANTAJE	DEZAVANTAJE
MODUL	- Uşor de calculat (nesemnificativ în prezent); - Poate fi utilizat pentru orice tip de scală; - Este singurul indicator pentru scale nominale;- - Corespunde unui scor real al distribuției;	- În general, nesigur, mai ales în cazul eşantioanelor mici, când se poate modifica dramatic la o modificare minoră a unei valori; - Poate fi greşit interpretat. Se identifică total cu un scor anume, fără a spune nimic despre celelalte valori; - Nu poate fi utilizat în statistici inferențiale;
MEDIANA	- Poate fi utilizată pe scale ordinale şi de interval\raport; - Poate fi utilizată şi pe distribuții de frecvență cu clase deschise sau scoruri nedeterminate la marginile distribuției;	- Poate să nu corespundă unei valori reale (N par); - Nu reflectă valorile distribuției (un scor extrem se poate modifica, fără a afecta Me); - Este mai puțin sigură în extrapolarea de la eşantion la populație; - Greu de utilizat în statistici avansate

Copyright © DEPARTAMENT ID 2008

MEDIA	- Reflectă valorile întregii distribuții; - Are multe proprietăți statistice dezirabile; - Adecvată pentru utilizare în statistici avansate;	- De obicei nu corespunde unei valori reale; - Nu este tocmai adecvată pentru scale ordinale; - Conduce la interpretări greşite pe distribuții asimetrice - Poate fi puternic afectată de scorurile extreme;

Sarcina de lucru nr. 2.3

Tabelul de mai jos conține două distribuții de valori (variabile). Una reprezintă scorurile la un test de evaluare a timidității, cealaltă, la un test de evaluare a sentimentului de singurătate.

Timiditate (1)	Singurătate (2)
29	27
28	35
36	30
41	51
25	30
15	20
33	47
40	42
33	40
20	33
35	28
26	40
32	22
23	15

Calculați şi scrieți care sunt, pentru fiecare dintre cele două variabile, următorii indicatori statistici:

(1). Mediana _____ Modul _____ Media _____

(2). Mediana _____ Modul _____ Media _____

Verificați răspunsurile corecte

Indicatori ai împrăştierii

Împrăştierea se referă la gradul de variabilitate a valorilor.

Indicatorii tendinței centrale se referă la ceea ce face ca valorile să se asemene, la caracteristica „comună” a valorilor unei distribuții. Indicatorii împrăştierii, de care vom vorbi în continuare, se referă la caracteristica de variabilitate, care descrie diferențele existente între valori. În cazul tendinței centrale este scoasă în evidență caracteristica valorilor unei distribuții de a

Copyright © DEPARTAMENT ID 2008

se „asemăna" unele cu altele, „asemănare" surprinsă de indicatorii tendinței centrale. În cazul împrăştierii, se urmăreşte descrierea tendinței valorilor de a se deosebi una de alta, de a se „sustrage" unei tendințe centrale prin îndepărtarea de aceasta.

Pentru evaluarea împrăştierii distribuțiilor statistice se utilizează mai mulți indicatori. Distingem două categorii de indicatori ai împrăştierii: *elementari* şi *sintetici*.

Principala caracteristică a indicatorilor elementari este aceea că surprind împrăştierea distribuției prin distanța dintre doar două valori ale acesteia.

Amplitudinea absolută (R de la Range)

Este dată de diferența dintre valoarea **maximă** şi valoarea **minimă** a unei distribuții

$R=X_{max}-X_{min}$ (formula 2.3)

Utilitatea ei este dată de faptul că ne indică în mod absolut plaja de valori între care se întinde distribuția.

Principalul dezavantaj constă în faptul că poate fi influențată de o singură valoare aflată la extremitatea distribuției.

Amplitudinea relativă

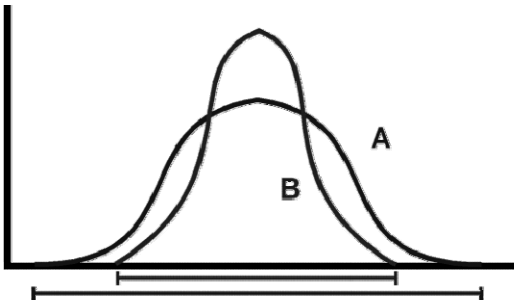
Este dată de raportul procentual dintre amplitudinea absolută şi media distribuției:

$$R\% = \frac{R}{m} * 100$$ (formula 2.4)

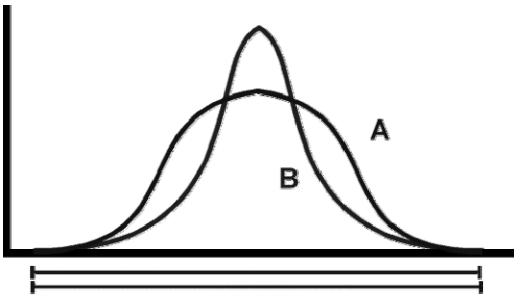
Este utilă atunci când cunoaştem plaja teoretică de variație a distribuției, putând astfel să facem o comparație cu plaja reală, obținută prin formula de mai sus.

Din cauză că amplitudinea utilizează doar cele două valori extreme ale distribuției, este un indicator imprecise al variabilității:

Exemple:



Distribuția A are o amplitudine mai mare dar şi o variabilitate mai mare decât distribuția B



Amplitudinea distribuțiilor A şi B sunt identice, dar distribuția A are mai multă variabilitate.

Copyright © DEPARTAMENT ID 2008

Abaterea quartilă (cvartilă, intercvartilă) (R_Q)

Quartilele (Q) sunt percentilele care împart distribuția în patru segmente egale. Ele sunt: Q1 (percentila 25); Q2 (percentila 50, sau *Me*); Q3 (percentila 75).

Abaterea quartilă este dată de diferența dintre valoarea corespunzătoare quartilei 3 şi valoarea corespunzătoare quartilei 1

$$R_Q = Q_3 - Q_1 \qquad \text{(formula 2.5)}$$

Nota bene: Se poate observa că este chiar distanța dintre limita superioară şi cea inferioară a casetei *Box-Plot* (valoarea H)

Abaterea semi-interquartilă (R_{SQ}):

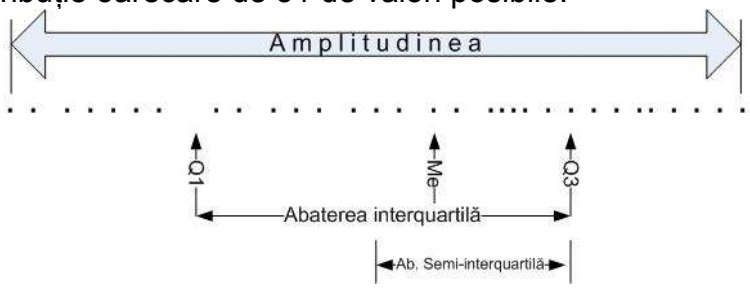
Semnifică distanța unui un scor „tipic" față de amplitudinea întregii distribuții şi se calculează ca media diferenței dintre quartila 3 şi quartila 1.

$$R_{SQ} = \frac{Q_3 - Q_1}{2} \qquad \text{(Formula 2.6)}$$

Într-o distribuție perfect simetrică $R_{SQ}=Q_2=Me$

R_{SQ} nu este afectată de valorile aberante fiind considerată, din acest motiv, un indicator „robust" al împrăştierii

O imagine de ansamblu a tipurilor de indicatori elementari ai împrăştierii ne este dată de figura de mai jos, unde am figurat prin puncte o distribuție oarecare de 31 de valori posibile.



Aşa cum am precizat, acest tip de indicatori ilustrează împrăştierea prin distanța dintre două puncte ale unei distribuții. Unul dintre avantajele lor este acela al uşurinței de calcul. Pe de altă parte, tocmai pentru că iau în seamă doar două dintre valorile distribuției, sunt vulnerabili şi nesiguri. Utilitatea lor este în general limitată dar sunt singurii care pot fi folosiți atunci când indicatorii sintetici (de care vom vorbi în continuare), nu pot fi calculați. Un alt dezavantaj al acestora este dificultatea de a fi utilizați în procedurile statistice avansate.

Spre deosebire de indicatorii elementari, indicatorii sintetici surprind împrăştierea unei distribuții prin luarea în considerarea abaterii fiecărei valori de la un anumit indicator al tendinței centrale. Cel mai uzual indicator de referință pentru împrăştiere este media. Aceasta pentru că, aşa cum ne amintim, media are avantajul de a fi o „concentrare" a tuturor valorilor unei distribuții.

Copyright © DEPARTAMENT ID 2008

Abaterea medie (d de la deviaţie medie)[2]

Distanța dintre o valoare anumită şi media distribuției se numeşte **abaterea valorii** (X_i-m). Dacă am dori să calculăm abaterea medie a unei distribuții nu ne-ar rămâne decât să însumăm abaterile individuale ale fiecărei valori şi să le împărțim la numărul acestora. Din păcate, media abaterilor într-o distribuție este întotdeauna egală cu zero (vezi proprietățile mediei). Acest fapt poate fi descris cu formula

$$\sum (X_i - m) / N = 0$$

unde X_i sunt valorile distribuției, *m* este media, iar N, numărul de valori.

X	**X_i – m**
5	(5 – 4.5) = .5
8	(8 – 4.5) = 3.5
3	(3 – 4.5) = -1.5
2	(2 – 4.5) = -2.5
5	(5 – 4.5) = .5
4	(4 – 4.5) = -.5
ΣX = 27	Σ(X_i-m) = 0
N = 6	
m = 4.5	

Aşa cum se observă în coloana „**X_i–m",** diferențele individuale însumate produc **Σ(X_i-m) = 0**. Acest lucru este valabil pentru orice fel de distribuție şi este una dintre proprietățile importante ale mediei.

Pentru a elimina acest inconvenient putem să luăm abaterile individuale în valoare absolută (fără semn).

X	**(X_i – m)**
5	(5 – 4.5) = 0.5
8	(8 – 4.5) = 3.5
3	(3 – 4.5) = 1.5
2	(2 – 4.5) = 2.5
5	(5 – 4.5) = 0.5
4	(4 – 4.5) = 0.5
ΣX = 27	Σ|X_i-m| = 9
N = 6	
m = 4.5	

Ca urmare, formula abaterii medii (***d***) poate fi scrisă astfel:

$$d = \frac{\sum | X_i - m |}{N} \qquad \text{(formula 2.7)}$$

[2] În continuare ne vom raporta la media de eşantionare. Se subînțelege că, pentru cazul unei populații, media va fi scrisă cu litera μ (miu).

Copyright © DEPARTAMENT ID 2008

Pentru cazul frecvențelor grupate, formula devine:

$$d = \frac{\sum |X_i - m| * f_i}{\sum f_i} \qquad \text{(formula 2.8)}$$

Abaterea medie este uşor de înțeles şi are semnificația de medie a distanțelor între fiecare scor şi media distribuției. Din păcate, nici ea nu este potrivită cu statisticile avansate

Dispersia (varianța, abaterea medie pătratică)

Notații uzuale:
s^2 (când se calculează pentru eşantion)
σ^2 (când se calculează pentru întreaga populație)
Pentru a elimina inconvenientul abaterilor de la medie de a avea suma egală cu zero, se operează ridicarea la pătrat a abaterilor valorilor individuale[3].

X	$(X_i - m)$	$(X_i - m)^2$
5	(5 – 4.5) = 0.5	0.25
8	(8 – 4.5) = 3.5	12.25
3	(3 – 4.5) = -1.5	2.25
2	(2 – 4.5) = -2.5	6.25
5	(5 – 4.5) = 0.5	0.25
4	(4 – 4.5) = -0.5	0.25
$\Sigma X = 27$	$\Sigma(X_i - m) = 0$	$\Sigma(X-m)^2 = 21.5$
N = 6		
m = 4.5		

Dacă însumăm abaterile ridicate la pătrat (pătratice) şi le împărțim la numărul valorilor, obținem *dispersia* (numită şi *varianță* sau *abatere medie pătratică*)

$$s^2 = \frac{\sum (X_i - m)^2}{N} \qquad \text{(formula 2.9)}$$

Notă: *Formula conține la numitor o anumită inexactitate care va fi discutată mai departe (vezi formula 2.11)*

Cu toate acestea, din cauza ridicării la pătrat, dispersia nu reprezintă o valoare foarte bună a împrăştierii (de ex., poate fi mai mare decât amplitudinea distribuției). Soluția acestui neajuns o constituie...

Abaterea standard

Abaterea standard este cel mai utilizat indicator al împrăştierii

Notații uzuale:
s (pentru eşantioane)
σ (pentru populație)
SD (*Standard Deviation*, în standardul APA)
ab.std.

[3] Această operație este permisă de proprietățile mediei

Copyright © DEPARTAMENT ID 2008

Abaterea standard se obține prin extragerea radicalului din expresia abaterii medii pătratice (dispersiei).

Formula de calcul: $s = \sqrt{\frac{\sum (X_i - m)^2}{N}}$ (formula 2.10)

NOTĂ: *Formula conține o inexactitate la numitor care va fi discutată mai departe (vezi formula 2.11)*

Pe datele din tabelul de mai sus: $s = \sqrt{\frac{21,5}{6}} = 1,89$

Operațiile succesive efectuate mai sus, ridicarea la pătrat şi extragerea radicalului, nu trebuie văzute ca operații artificiale, „gratuite". Aceste operații nu se referă la valorile distribuției ci la abaterile de la medie, ceea ce conduce la rezultate diferite care exprimă, într-o altă formă, aceeaşi caracteristică de împrăştiere a valorilor originale.

Corecția indicatorilor împrăştierii calculați pentru eşantioane

Formulele 2.8 şi 2.9 au la numitor valoarea N (volumul eşantionului). Fără a intra în detalii, vom spune că valorile astfel calculate, ale dispersiei şi abaterii standard, pentru un eşantion, conțin o imprecizie (*bias*) care conduce la subestimarea împrăştierea la nivelul populației. Chiar dacă luăm în considerare un număr mare de eşantioane, extrase succesiv dintr-o anumită populație, indicatorii împrăştierii vor fi mai mici decât împrăştierea la nivelul întregii populații.

Corecția se face prin utilizarea la numitor a expresiei *N-1*. În acest mod, cu cât eşantionul este mai mic, cu atât indicatorul respectiv al împrăştierii va fi influențat mai mult de expresia de la numitor.

Expresia *N-1* poartă numele de „*grade de libertate*". Pentru a-i înțelege semnificația, este bine să ne gândim la faptul că, într-o distribuție de 3 valori (de exemplu: 1,3,8) media este 4, iar abaterile de la medie sunt –3, -1, 4. Suma lor este zero. Ca urmare, este suficient să cunoaştem cel puțin două din cele trei valori pentru a o afla pe a treia. Altfel spus, doar două valori sunt libere să se modifice, a treia (ultima) fiind determinată de acestea.

Formulele corecte devin astfel:

Dispersia: $s^2 = \frac{\sum (X_i - m)^2}{N - 1}$ (formula 2.11)

Abaterea standard: $s = \sqrt{\frac{\sum (X_i - m)^2}{N - 1}}$ (formula 2.12)

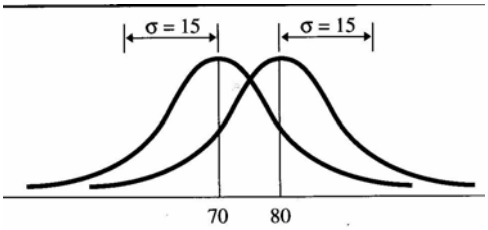
Formulele inițiale, de definiție, rămân corecte pentru situația în care se urmăreşte doar descrierea caracteristicii de împrăştiere pentru eşantionul respectiv. Atunci când se urmăreşte însă extrapolarea acestei valori la nivelul populației, utilizarea formulei corectate este absolut necesară.

Copyright © DEPARTAMENT ID 2008

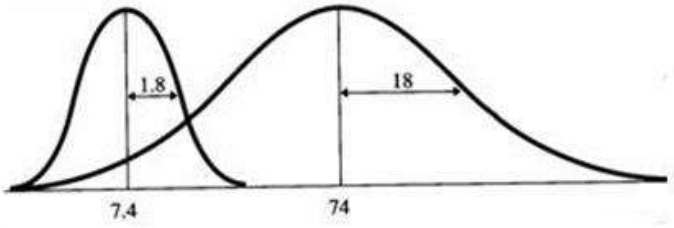
Proprietățile abaterii standard

Abaterea standard este, aşa cum vom vedea, indicatorul principal al împrăştierii utilizat în diverse proceduri statistice avansate. Pentru a-i justifica modul de utilizare în diverse formule, trebuie să reținem câteva proprietăți fundamentale ale abaterii standard:

1. Dacă se adaugă/scade o constantă la fiecare valoare a unei distribuții, abaterea standard nu este afectată



2. Dacă se multiplică/divide fiecare valoare a unei distribuții cu o constantă, abaterea standard se multiplică/divide cu acea constantă



3. Abaterea standard față de medie este mai mică decât abaterea standard față de orice altă valoare a unei distribuții

Coeficientul de variație

Abaterea medie şi abaterea standard se exprimă în unitățile de măsură ale variabilei de referință. De exemplu, pentru o distribuție de timpi de reacție, exprimați în sutimi de secundă, s=2.14 înseamnă că împrăştierea standard este de 2.14 sutimi de secundă.

Dacă acelaşi eşantion face şi un test de coordonare a mişcărilor, evaluat în număr de „ieşiri din traseu" a căror abatere standard este s=20.94, nu putem compara omogenitatea celor două serii de valori. Adică, nu putem spune dacă eşantionul este mai omogen sau mai puțin omogen din perspectiva unei dintre cele două performanțe.

Dintre soluțiile posibile pentru eliminarea acestui neajuns, cea mai des utilizată este coeficientul de variație (variabilitate), notat cu *cv* (sau *v*), propus de Pearson. Se calculează ca raport între abaterea standard şi medie. Poate fi exprimat şi procentual conform formulei de mai jos:

$$cv = \frac{s}{m} * 100 \qquad \text{(formula 2.13)}$$

Valoarea acestui coeficient exprimă un raport procentual dintre abaterea standard şi medie. Cu cât este mai mare, cu atât media putem spune că media este mai puțin „reprezentativă" pentru distribuția

Copyright © DEPARTAMENT ID 2008

respectivă, dată fiind ponderea ridicată a împrăştierii. Utilizarea coeficientului de variație este limitată la valorile măsurate pe scale de raport, cu origine naturală 0. În cazul a două variabile a căror origine este diferită una de alta, diferențele dintre valori (abaterea standard) rămân aceleaşi dar media se schimbă, fapt care face ca raportul exprimat în formulă să fie modificat iar comparația a doi coeficienți de variație, irelevantă. În plus, pe o scală de interval cu valori negative se poate ajunge la medie egală cu 0, ceea ce face formula inaplicabilă.

Utilitatea coeficientului de variație vine de la faptul că valoarea sa mai este legată de unitatea de măsură. Diferența dintre două valori *cv* poate fi interpretată ca diferență de împrăştiere a celor două variabile, chiar dacă măsoară lucruri diferite.

Sunt propuse anumite limite de interpretare a acestui indicator, astfel:

- dacă *cv*<15%, împrăştierea este mică şi, deci, media este reprezentativă
- dacă *cv* este cuprins între 15%-30%, împrăştierea este mijlocie şi media este suficient de reprezentativă
- dacă *cv* este mai mare de 30%, împrăştierea este mare şi media are o reprezentativitate redusă

Calcularea coeficientului de variație a unei distribuții, înainte de integrarea ei în proceduri statistice inferențiale, este o metodă utilă de verificare a măsurii în care media, pe care se bazează de cele mai multe ori procedurile inferențiale, este legitimă.

Alegerea indicatorului împrăştierii

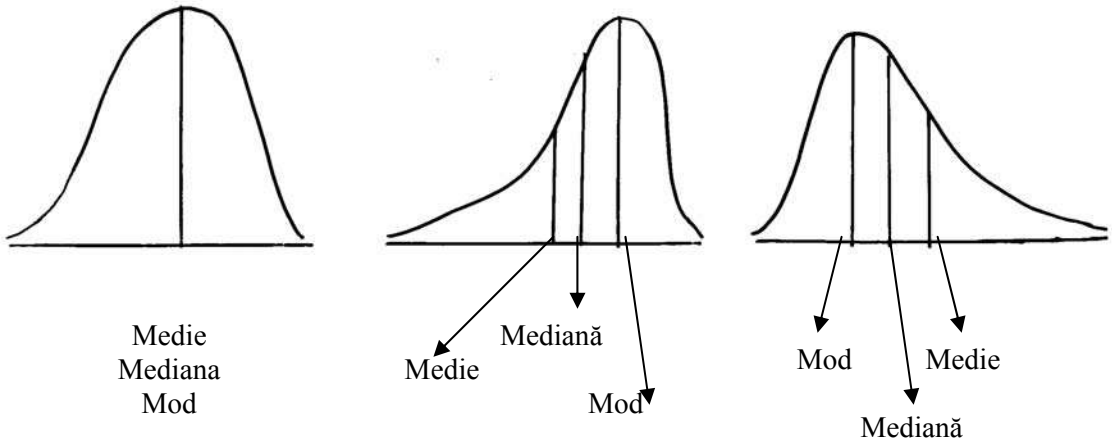
- Abaterea standard este cea mai utilizată pentru scale de măsurare interval/raport. Realizează cea mai bună combinație între calitatea estimării şi posibilitatea de a fundamenta inferențe statistice.
- Amplitudinea este un indicator nesigur şi care nici nu poate fi calculat în cazul scalelor nominale
- Pe distribuții cu valori nedeterminate sau cu intervale deschise, se alege abaterea interquartilă (semi-interquartilă).

Indicatori ai formei distribuției

Expresia grafică a distribuțiilor poate fi descrisă sub două aspecte esențiale: simetria şi boltirea. O distribuție este <u>simetrică</u> atunci când valorile acesteia se împart în mod egal de o parte şi de alta a valorilor tendinței centrale. Se numesc asimetrice (*skewed*) distribuțiile ale căror valori se concentrează fie în zona valorilor mici (spre stânga) fie în zona valorilor mari (spre dreapta).

Copyright © DEPARTAMENT ID 2008

Distribuție: simetrică asimetrică negativ asimetrică pozitiv



Figurile de mai sus arată cum se plasează cei trei indicatori ai tendinței centrale în funcție de simetria distribuției:

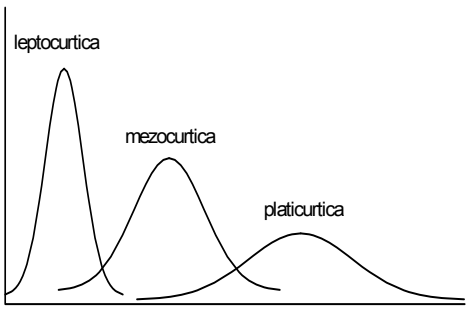
- În cazul distribuțiilor (perfect) simetrice, **Mo**, **Me** şi **m** se plasează pe aceeaşi valoare
- În cazul distribuțiilor asimetrice cei trei indicatori au poziții diferite (vezi figura).
- Mediana se plasează întotdeauna între mod şi medie. Din acest motiv, mediana este cea mai reprezentativă valoare pentru distribuțiile asimetrice
- Media este afectată de valorile extreme, cu atât mai mult cu acestea sunt mai puternic deviate. Ca urmare, în cazul distribuțiilor puternic asimetrice, media nu este un indicator veridic al tendinței centrale.

Descrierea numerică a caracteristicii de simetrie/asimetrie se face cu ajutorul unui indicator statistic specific, numit indicator de „simetrie" sau de „oblicitate" (*skewness*, în limba engleză).

Pentru o curbă absolut simetrică, indicele de oblicitate (*skewness*) are valoarea 0 (zero), primind valori pozitive pentru curbele asimetric pozitive şi valori negative pentru cele asimetric negative. Ca reper general de apreciere, recomandat de cei mai mulți autori, un indice de oblicitate a cărui valoare depăşeşte +1/-1 semnalează o asimetrie pronunțată a distribuției.

Caracteristica de boltire (*kurtosis*, în terminologia engleză) indică gradul de extindere pe verticală a curbelor de distribuție. În termeni generali, sub aspectul boltirii, curbele pot fi de trei categorii:

- Leptokurtice, cu majoritatea valorilor distribuite în zona mediei (au o formă „înaltă" şi „subțire")
- Mezokurtice, cu o prezență „moderată" a valorilor în zona mediei
- Platikurtice, cu valori medii relativ puține şi o formă aplatizată



Copyright © DEPARTAMENT ID 2008

Desigur, o curbă poate fi în acelaşi timp şi asimetrică şi boltită excesiv, chiar dacă imaginea de mai sus ilustrează boltirea pe curbe simetrice.

Indicatorul numeric al boltirii (*kurtosis*) are o plajă de variație în jurul valorii zero (care înseamnă boltire medie, „normală", *mezocurtică*). Indicele de boltire pozitivă indică o curbă „înaltă" (*leptocurtică*), iar indicele de boltire negativă, o curbă „aplatizată" (*platicurtică*). La fel ca şi în cazul indicelui de oblicitate (*skewness*), cu cât acesta este mai îndepărtat de valorile +1/-1, avem de a face cu distribuții cu abatere accentuată de la boltirea „normală".

Sarcina de lucru nr. 2.4

Pentru cele două variabile de la sarcina de lucru nr 2.3 („timiditate" şi „singurătate"), calculați şi scrieți valorile cerute mai jos:

(1) amplitudinea _____ abaterea quartilă _____ abaterea semiinterquartilă ____

abaterea medie pătratică _____ abaterea standard _____ coeficientul de variație

_____

(2) amplitudinea _____ abaterea quartilă _____ abaterea semiinterquartilă ____

abaterea medie pătratică _____ abaterea standard _____ coeficientul de variație

_____

Verificați răspunsurile corecte

Mai jos, încercuiți răspunsul şi apoi verificați răspunsul corect

3. Care dintre indicatorii împrăştierii (amplitudine, abatere interquartilă, abatere standard) ar trebui aleşi pentru fiecare dintre următoarele situații:

a) Distribuția este puternic asimetrică, având câteva valori extreme într-o singură direcție a curbei
b) Intenționați să utilizați proceduri statistice avansate (de exemplu, să emiteți aprecieri asupra populației pe baza datelor de eşantion)
c) Vreți să ştiți întinderea maximă a unei distribuții
d) Vreți ca fiecare valoare a distribuției să fie luată în considerare
e) Valoarea cea mai mare a distribuției este „*mai mult de 10*"

Copyright © DEPARTAMENT ID 2008

Valori extreme ale distribuției

Valorile extreme reprezintă valori excesive ale unei distribuții. Identificarea lor este necesară pentru a evita efectul pe care îl au asupra valorilor tendinței centrale, în primul rând asupra mediei. Una dintre metodele de identificare este analiza grafică de tip *Box-and-Whisker-Plot* (pe scurt *Box-Plot*), elaborată de Tukey.

În esență, reprezentarea *Box-Plot* (vezi imaginea) este constituită dintr-o casetă (dreptunghi), a cărui limită inferioară este plasată în dreptul percentilei 25, limita superioară fiind plasată în dreptul percentilei 75. Cu alte cuvinte, caseta cuprinde 50% dintre valorile unei distribuții. Distanța dintre valorile limită ale casetei se numeşte H.

Linia din interiorul casetei marchează valoarea mediană (Me)

„Mustățile" care pornesc de la limita superioară şi inferioară a casetei, au o lungime maximă egală cu 1,5 H. În acel punct se plasează ultima valoare „legitimă" a distribuției. Orice valoare mai mică sau mai mare de acestea, sunt definite ca extreme (*Outliers*)

Un exemplu de creare a reprezentării box plot: Vom utiliza distribuția scorurilor QI prezentată anterior, la care am adăugat două valori suplimentare (135 şi 142), alese intenționat pentru a fi mai mari decât restul valorilor.

101	94	87	117	115	116	91	113	96	105	135
92	107	118	114	98	112	101	114	107	109	142
97	109	124	102	118	113	116	106	108	89	
106	108	115	92	97	102	108	102	109	114	
107	104	110	101	101	121	125	86	109	123	

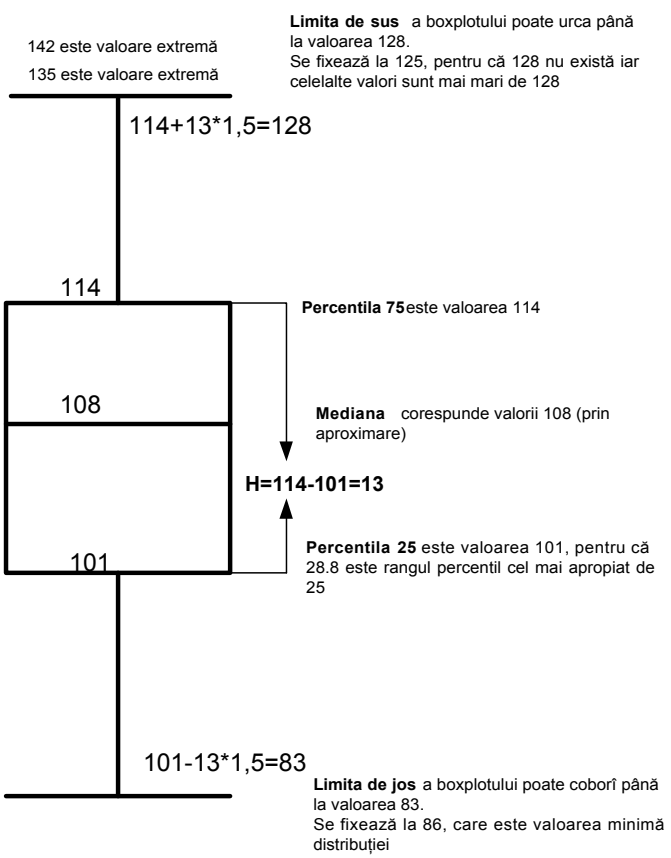
Pentru a face reprezentarea box plot facem mai întâi tabela de frecvențe simple, cu scopul calculării percentilelor. Tabelul de frecvențe alăturat cuprinde valorile ordonate ale distribuției, între de la valoarea cea mai mică (86) şi se cea mai mare (142). Pe coloana **frc%** se află frecvențele cumulate procentuale (percentilele). Pentru box plot identificăm percentilele 25 şi 75. Ele corespund valorilor 101 (este valoarea cea mai apropiată de 25 pe coloana **frc%**) şi, respectiv, 114. Am obținut astfel, limita inferioară şi superioară a casetei. Mediana (percentila 50) corespunde valorii 108 (frc%=53.8, prin aproximare). Diferența dintre valorile corespunzătoare percentilelor 25 şi 50 este 13 (114-101). Astfel putem determina limitele prelungirilor superioară şi inferioară ale casetei care sunt: 114+13*1.5=128 (aproximare) pentru prelungirea superioară şi, respectiv 101-13*1,5=83 (aproximare) pentru cea de jos. Am obținut astfel toate valorile necesare trasării *box plotului*.

Imaginea de mai jos prezintă tabelul distribuției şi *boxplot*-ul corespunzător[4]:

[4] În mod normal, reprezentarea boxplot se construieşte independent de tabelul de frecvențe. Dacă le-am asociat în imagine, am făcut-o doar cu scop didactic, pentru a pune mai clar în evidență mecanismul de elaborare.

Copyright © DEPARTAMENT ID 2008

Valori QI	fa	fr%	frc% (rang percentil)
142	1	1,9	100,0
135	1	1,9	98,1
125	1	1,9	96,2
124	1	1,9	94,2
123	1	1,9	92,3
121	1	1,9	90,4
118	2	3,8	88,5
117	1	1,9	84,6
116	2	3,8	82,7
115	2	3,8	78,8
114	3	5,8	75,0
113	2	3,8	69,2
112	1	1,9	65,4
110	1	1,9	63,5
109	4	7,7	61,5
108	3	5,8	53,8
107	3	5,8	48,1
106	2	3,8	42,3
105	1	1,9	38,5
104	1	1,9	36,5
102	3	5,8	34,6
101	4	7,7	28,8
98	1	1,9	21,2
97	2	3,8	19,2
96	1	1,9	15,4
94	1	1,9	13,5
92	2	3,8	11,5
91	1	1,9	7,7
89	1	1,9	5,8
87	1	1,9	3,8
86	1	1,9	1,9
Total	**52**	**100,0**	


142 este valoare extremă
135 este valoare extremă
Limita de sus a boxplotului poate urca până la valoarea 128.
Se fixează la 125, pentru că 128 nu există iar celelalte valori sunt mai mari de 128
114+13*1,5=128
114
Percentila 75 este valoarea 114
108
Mediana corespunde valorii 108 (prin aproximare)
H=114-101=13
Percentila 25 este valoarea 101, pentru că 28.8 este rangul percentil cel mai apropiat de 25
101
101-13*1,5=83
Limita de jos a boxplotului poate coborî până la valoarea 83.
Se fixează la 86, care este valoarea minimă distribuției

Tratarea valorilor extreme

Punerea în evidență a unor valori extreme ridică problema modului lor de tratare a acestor valori. În acest scop, trebuie să avem în vedere două aspecte:

1. Stabilirea naturii valorilor extreme, care pot apare în următarele situații:
   - ✓ erori de înregistrare (tastare);
   - ✓ erori de măsurare;
   - ✓ rezultate influențate de anomalii ale condițiilor experimentale;
   - ✓ eşantionul a fost extras dintr-o populație asimetrică;
   - ✓ valorile respective fac parte din altă populație de valori
   - ✓ eşantion prea mic.
2. Tratarea lor pe una din căile posibile:
   - ✓ eliminare (dacă sunt erori necorectabile);
   - ✓ corectare (dacă este posibil);
   - ✓ utilizarea ***mediei 5%trim,*** adică a mediei care nu ține cont de 5% din numărul valorilor de la fiecare din cele două extremități ale distribuției;
   - ✓ transformare (dacă datele sunt corecte şi, totuşi, dorim să evităm efectul lor asupra indicatorilor sintetici);
     - o există diverse metode de transformare: extragerea radicalului din toate valorile distribuției, logaritmarea distribuției

Analiza valorile extreme reprezintă unul dintre obiectivele principale ale fazelor preliminare de analiză a datelor. Prezența lor este de natură să

Copyright © DEPARTAMENT ID 2008

aibă efecte majore asupra rezultatelor fapt care trebuie luat în considerare la alegerea procedurilor statistice inferențiale.

Rezumatul unității de învățare

- Statistica descriptivă are drept obiective organizarea, sintetizarea şi descrierea datelor.
- Tehnicile statisticii descriptive sunt globale sau sintetice
- Statisticile descriptive globale sunt numerice (analiza de frecvențe simple şi grupate) şi grafice.
- Rangul percentil se defineşte ca procentajul datelor valorilor dintr-o distribuție care se află până la o anumită valoare inclusiv.
- Percentila este valoarea dintr-o distribuție care corespunde unui anumit rang percentil.
- Un indicator statistic concentrează într-o singură valoare o anumită caracteristică a distribuției
- Statisticile descriptive sintetice sunt reprezentate de indicatorii tendinței centrale (modul, mediana, media), indicatorii împrăştierii sau variabilității (amplitudine, abatere interquartilă, abaterea medie, dispersia, abaterea standard) şi indicatorii formei distribuției (simetrie şi boltire).
- Cei mai frecvent utilizați indicatori statistici sunt media şi abaterea standard.

Răspunsuri corecte la sarcinile de lucru

Sarcina de lucru nr. 2.1

1. a
2. c
3. a (50%)
4. O valoare care exprimă raportul dintre frecvența unei valori şi 1
5. Valoarea respectivă apare în 7% din totalul valorilor unei distribuții
6. Trebuie să fie multiplu al mărimii intervalului de grupare ales
7. între 5 şi15
8. Ranguri percentile
9. Percentilă

Sarcina de lucru nr. 2.2

1. variabile măsurate pe scale de interval/raport

Copyright © DEPARTAMENT ID 2008

2. ilustrează nu doar forma distribuției ci şi valorile din care este compusă
3. Atunci când suma valorilor reprezentate are semnificația unui „întreg”
4.

Stem	Leaf
1	5
2	0,3,5,6,8,9
3	2,3,3,5,6
4	0,1

Sarcina de lucru nr. 2.3

Variabila (1): modul=33; mediana=0.5; media=29.7
Variabila (2): modul=30 şi 40 ; mediana=31.5; media=32.8

Precizări:
Variabila (2) este multimodală, 30 este modul cel mai mic.

Sarcina de lucru nr. 2.4

Pentru cele două variabile de la sarcina de lucru nr 2.3 („timiditate” şi „singurătate”), calculați şi scrieți valorile cerute mai jos:

(1) amplitudinea=26; abaterea quartilă=10.7; abaterea semiinterquartilă=5.35; abaterea medie pătratică=55.6; abaterea standard=7.4; coeficientul de variație=24.9%;

(2) amplitudinea=36; abaterea quartilă=14.7; abaterea semiinterquartilă=7.35; abaterea medie pătratică=107,33; abaterea standard=10.36; coeficientul de variație=31.5%;

3. Se utilizează următorii indicatori:
   f) abatere interquartilă sau semiinterquartilă
   g) abatere standard
   h) amplitudine
   i) abaterea standard
   j) abatere interquartilă sau semiinterquartilă

Lucrarea de evaluare nr. 1.2

Lucrarea de evaluare va fi publicată pe portal (http://portal.credis.ro). Data limită de trimitere este preziua tutorialului. După acest termen lucrările nu mai sunt acceptate.

Copyright © DEPARTAMENT ID 2008

Bibliografie minimală

Bibliografia de bază

- Marian Popa, (2008), Statistică pentru psihologie. Teorie şi aplicații SPSS, Polirom
- Pagina web a cursului, la adresa: ***www.mpopa.ro***

Bibliografie suplimentară

- Clocotici, V., & Stan, A. (2000). Statistica aplicata in psihologie. Iasi: Polirom, p. 63-73
- Rotaru, T. (coord.). (1999). Metode statistice aplicate in stiintele sociale. Iasi: Polirom. p. 42-61
- Vasilescu, I. P. (1992). Statistica informatizata pentru stiinte despre om (Vol. 1-2). Bucuresti: Editura militara., p.95-116

Copyright © DEPARTAMENT ID 2008

UNIVERSITATEA DIN BUCUREŞTI

FACULTATEA DE PSIHOLOGIE ŞI ŞTIINȚELE EDUCAȚIEI

DEPARTAMENTUL DE ÎNVĂȚĂMÂNT LA DISTANȚĂ

STATISTICĂ PSIHOLOGICĂ ŞI PRELUCRAREA COMPUTERIZATĂ A DATELOR

(Modulul II)

Statistică inferențială, noțiuni de bază
Teste statistice parametrice

Conf. Univ. dr. Marian Popa

Universitatea din Bucureşti
Editura CREDIS
2008

Copyright © DEPARTAMENT ID 2008

CUPRINS

STATISTICĂ INFERENȚIALĂ, NOȚIUNI DE BAZĂ 5
Obiectivele unității de învățare 5
Scoruri standard 5
Alte tipuri de scoruri standardizate 6
Distribuția normală (Gauss) 7
Proprietățile distribuției normale 7
Distribuția normală z 8
Aria de sub curba normală văzută ca probabilitate 10
Distribuții reale şi distribuții normale z 11
Distribuția de eşantionare 11
Populație şi eşantion 11
Reprezentativitatea eşantionului 12
Distribuția mediei de eşantionare 13
Împrăştierea distribuției de eşantionare (eroarea standard a mediei) 14
Teorema limitei centrale 15
Sarcina de lucru nr. 2. 1 16
Notele standardizate z pentru eşantioane (grupuri) 16
Ipoteze şi decizii statistice 17
Ipoteza cercetării 17
Ipoteza statistică (de nul) 18
Distribuția ipotezei de nul 18
Testul *z* pentru un singur eşantion 19
Procedura de calcul 19
Decizia statistică 20
Decizii statistice unilaterale şi bilaterale 21
Estimarea intervalului de încredere pentru media populației 22
Testul t (Student) pentru un singur eşantion 23
Publicarea rezultatelor testului z sau t 24
Sarcină de lucru nr. 2. 2 25
Erori statistice; Puterea testului statistic; Mărimea efectului 26
Erori statistice 26
Eroarea de tip I 27
Eroarea de tip II 27
Eroarea de tip III 28
Puterea testului 29
Factori care contribuie la creşterea puterii testelor statistice 29

Copyright © DEPARTAMENT ID 2008

Mărimea efectului *31*
Calcularea mărimii efectului pentru testul *z* (*t*) pentru un singur eşantion 32
Relația dintre mărimea efectului şi puterea testului 33

Interpretare rezultatului unui test statistic *33*

Rezumatul unității de învățare **34**

Răspunsuri la sarcinile de lucru *35*

Lucrarea de evaluare nr. 2.1 ..... **36**

TESTE STATISTICE PARAMETRICE **37**

Obiectivele unității de învățare şi informații introductive **37**

Testarea diferenței dintre mediile a două eşantioane independente **37**

Distribuția ipotezei de nul pentru diferența dintre medii independente *37*

Procedura statistică pentru testarea semnificației diferenței dintre mediile a două eşantioane *38*
a. Testul *t* pentru dispersii diferite 40
b. Testul *t* pentru dispersia cumulată 40
Mărimea efectului 42
Limitele de încredere ale diferenței dintre medii 43
Interpretarea rezultatului la testul *t* pentru eşantioane independente 44
Publicarea rezultatului 44
Condițiile în care putem calcula testul *t* pentru eşantioane independente 44
Când se utilizează testul *t* pentru eşantioane independente? 44

Sarcina de lucru nr. 2.4 ..... *45*

Analiza de varianță (mai mult de două eşantioane independente) **45**

Cadrul conceptual pentru analiza de varianță unifactorială *46*

Fundamentarea procedurii de calcul ANOVA *48*

Interpretarea raportului F *49*
Distribuția Fisher 50
Mărimea efectului pentru testul F 52

Analiza „post-hoc” *54*
Publicarea rezultatului testului F (ANOVA) 55
Avantajele ANOVA 55
Condiții pentru utilizarea testului ANOVA 55

Sarcina de lucru 2.5 ..... *56*

Testul t pentru diferența dintre medii pentru eşantioane dependente **56**

Sarcina de lucru nr. 2.6 ..... *61*

Testarea asocierii dintre două variabile măsurate pe aceiaşi subiecți **62**

Coeficientul de corelație liniară (Pearson) *62*
Corelația liniară 63
Reprezentarea grafică a corelației 63
Calcularea coeficientului de corelație liniară Pearson 64
Corelație şi cauzalitate 66
Natura liniară a corelației Pearson 66
Mărimea efectului coeficientului de crelație 68
Coeficientul de determinare 69
Limitele de încredere pentru coeficientul de corelație *r* 70
Semnificația diferenței dintre doi coeficienți de corelație 71

Copyright © DEPARTAMENT ID 2008

Condiții pentru calcularea coeficientului de corelație Pearson....72
Utilizarea coeficientul de corelație....72
Publicarea rezultatului corelației....72
Sarcina de lucru 2.7....*72*
Rezumatul unității de învățare....73
Răspunsuri la sarcinile de lucru....74
Sarcina de lucru 2.3....74
Lucrarea de evaluare nr. 2.2....75
Bibliografie minimală....75
ANEXE – TABELE STATISTICE....76
Anexa 1. Tabelul distribuției valorilor sub curba normală z....76
Anexa 2. Tabelul valorilor critice pentru distribuția t Student (unilateral)....78
Anexa 3. Tabelul parțial al distribuției F pentru α=0.05....79
Anexa 5. Tabelul Fisher de transformare a valorilor r în scoruri Z....81

Copyright © DEPARTAMENT ID 2008

STATISTICĂ INFERENȚIALĂ, NOȚIUNI DE BAZĂ

Obiectivele unității de învățare

Parcurgerea acestei unități, va permite studenților:

- ✓ să calculeze scoruri standard corespunzătoare unor scoruri brute
- ✓ să definească proprietățile scorurilor standardizate z
- ✓ să definească proprietățile distribuției normale
- ✓ să definească caracteristicile distribuției de eşantionare
- ✓ să definească ipoteza cercetării şi ipoteza de nul

Scoruri standard

Modalitatea de a exprima semnificația unei anumite valori dintr-o distribuție prin raportare la parametrii distribuției (medie şi abatere standard) este **scorul standardizat** *z* (numit şi notă *z* sau scor *z*)**.** Aceasta măsoară distanța dintre o anumită valoare şi media distribuției, în abateri standard:

$$z = \frac{x - m}{s}$$

unde x reprezintă oricare dintre valorile distribuției

Pentru cele două distribuții de mai sus, scorurile *z* se calculează astfel:

$$z_I = \frac{70-60}{5} = +2.0 \qquad \text{respectiv,} \qquad z_{II} = \frac{70-60}{20} = +0.5$$

Iar în cazul în care pentru distribuția II am avea un scor de 45:

$$z_{II} = \frac{45-60}{20} = -0.75$$

Semnul „–„ la rezultat ne arată că performanța este mai mică decât media, mai precis, se află la 0.75 abateri standard sub medie. Semnul „+" indică o valoare standardizată peste medie, indicând, în exemplul de mai sus, că se plasează la o jumătate de abatere standard deasupra mediei.

Scorul *z* se numeşte „**scor standardizat**" (notă standardizată), deoarece exprimă distanța unei valori față de media distribuției din care face parte în unități ale abaterii standard. De aici decurge unul din avantajele lui importante, acela de a putea fi utilizat pentru a compara valori care provin din distribuții diferite, indiferent de unitatea de măsură a fiecăreia.

Exemplu: Dacă un subiect obține un scor echivalent cu *z*=+0.2 la un test de calcul aritmetic şi un scor echivalent cu *z*=+0.1, la un test de reprezentare spațială, se poate spune că are o performanță mai bună la primul test decât la al doilea.

Calcularea valorii atunci când cunoaştem parametrii scorului *z*

Copyright © DEPARTAMENT ID 2008

Dacă am calcula scorurile (notele) *z* pentru fiecare dintre valorile unei distribuții, am obține o „distribuție în scoruri *z*" a acelei distribuții. În tabelul următor, distribuția X a fost transformată în distribuție *z*.

X	*z*
14	+0.50
11	-0.75
10	-1.17
16	+1.34
13	+0.08
N=5 $\sum X = 64$ m=12.8 s=2.38	N=5 $\sum Z = 0$ m=0 s=1

Utilizând proprietățile de transformare a formulei de definiție a scorului *z*, putem calcula o anumită valoare atunci când cunoaştem valoarea lui *z* şi parametrii distribuției, astfel:

Dacă $z = \frac{x - m}{s}$ atunci:

x=*z**s+m adică, pentru ultimul exemplu, x=-0,75*2.38+12.8=11

Proprietățile scorurilor *z*

1. Media unei distribuții *z* este întotdeauna egală cu 0. Aceasta rezultă din proprietatea mediei de a se diminua corespunzător dacă se extrage o constantă din fiecare valoare a unei distribuții. Formula de calcul pentru *z* implică scăderea unei constante din fiecare valoare a distribuției. Aceasta înseamnă că şi media noii distribuții (*z*) se va reduce cu constanta respectivă. Dar această constantă este însăşi media distribuției originale, ceea ce înseamnă că distribuția *z* va avea media egală cu zero, ca rezultat al diminuării mediei cu ea însăşi.

2. Abaterea standard a unei distribuții *z* este întotdeauna 1. Acest fapt decurge prin efectul cumulat al proprietăților abaterii standard. Prima proprietate afirmă că în cazul scăderii unei constante (în cazul scorurilor *z*, media) din valorile unei distribuții, abaterea standard a acesteia nu se modifică. A doua proprietate afirmă că în cazul împărțirii valorilor unei distribuții la o constantă, noua abatere standard este rezultatul raportului dintre vechea abatere standard şi constantă. Dar constanta de care vorbim este, în cazul distribuției *z*, chiar abaterea standard. Ca urmare, noua abatere standard este un raport dintre două valori identice al cărui rezultat, evident, este 1.

Alte tipuri de scoruri standardizate

Scorurile **z** prezintă un avantaj important, permit compararea valorilor unei distribuții şi a valorilor provenind din distribuții diferite, ca urmare a faptului ca se exprimă în abateri standard de la medie. Totuşi se impune o anumită precauție în comparația pe baza scorurilor *z* atunci când distribuțiile au forme diferite şi, mai ales, asimetrii opuse.

Notele *z* au însă şi unele dezavantaje: se exprimă prin numere mici, cu zecimale, (greu de manipulat intuitiv) şi, în plus, pot lua valori negative. Aceste dezavantaje pot fi uşor înlăturate printr-un artificiu de calcul care să conducă la note standardizate convenabile, ce corespund anumitor nevoi practice specifice. În tabelul de mai jos sunt descrise câteva tipuri de note standard calculate pe baza notelor *z*.

Copyright © DEPARTAMENT ID 2008

	Formula bazată pe „z”	Formula desfășurată	m	s
Note z			**0**	**1**
Note T (Thurstone)[1]	**50+10*z**		**50**	**10**
Note H (Hull)	**50+14*z**		**50**	**14**
QI (Binet)	**100+16*z**		**100**	**16**
QI (Wechsler)	**100+15*z**		**100**	**15**
SAT (Scholastic Assessment Test)	**500+100*z**		**500**	**100**

Observații:

- Toate variantele sunt obținute prin transformarea operată pe distribuția de note *z*.
- La nici una dintre variante nu mai avem valori negative (cu condiția ca distribuția să nu aibă o variabilitatea aberantă).
- Zecimalele nu mai sunt semnificative (ele rezultă din calcule, dar sunt ignorate).
- Distribuțiile variantelor oscilează în jurul unei valori medii specifice, sub care se află 50% din valori, şi peste care se află restul de 50% dintre valori.
- Scorurile standard mari indică valori mari, iar scorurile standard mici indică valori mici. Acest fapt poate crea dificultăți în unele cazuri. Să luăm următorul exemplu: Un subiect realizează 145 răspunsuri corecte la un test de calcul aritmetic (m=120, s=12) şi un timp de reacție de 0.15 sec., la un test de reactivitate (m=0,11, s=0,05). În acest caz, notele T corespunzătoare celor două performanțe sunt: T_1=50+10*(145-120)/12=**70**, respectiv T_2=50+10*(0,15-0,11)/0,05=**58**. Cu alte cuvinte, ar rezulta că la ambele teste subiectul nostru a obținut un rezultat peste medie. Dar această concluzie este falsă, dacă ținem cont că la testul de reactivitate un timp mai mare înseamnă o performanță mai scăzută. Soluția problemei constă în modificarea semnului expresiei de calcul, în funcție de semnificația calitativă a valorilor distribuției. În acest mod, rezultatul transformării în notă standard la testul de reactivitate devine: T_2=50-10*(0,15-0,11)/0,05=**42**, ceea ce indică exact semnificația de performanță sub medie. Raportată la valoarea medie a distribuției T, scorul 58 este echivalent cu 42, sub aspectul distanței față de medie (8 unități). Diferența constă în faptul că valoarea 42 exprimă şi în mod intuitiv, nu doar cantitativ, evoluția performanței la test. O asemenea transformare nu este obligatorie, se poate utiliza oricare dintre formule, cu semnul plus, sau minus. În orice caz, trebuie să precizăm semnificația valorilor mari si mici pentru distribuțiile cu care operăm.

Distribuția normală (Gauss)

Proprietățile distribuției normale

Reprezentarea grafică a rezultatelor măsurărilor reale poate lua diverse forme, curba distribuției putând fi unimodală sau multimodală, aplatizată sau înaltă, simetrică sau asimetrică. În statistică există însă un tip special de distribuție, numită „distribuție normală”, care corespunde reprezentării grafice a unei caracteristici pentru care există un mare număr de măsurări, tinzând spre infinit. Această distribuție este

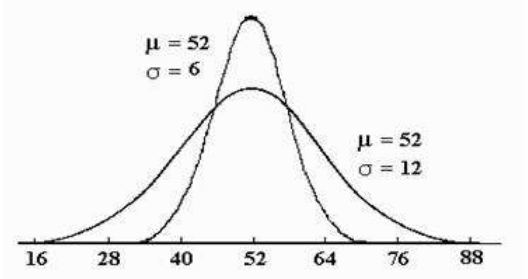
[1] Cunoscute mai ales datorită utilizării în evaluarea scalelor la MMPI, unul dintre cele mai celebre teste de personalitate.

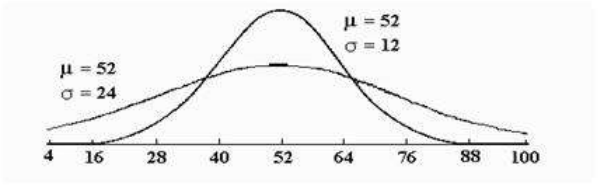
Copyright © DEPARTAMENT ID 2008

numită „teoretică" pentru că nu este rezultatul unui proces real de măsurare, ci reprezintă un model teoretic. Conceptul de „curbă normală" (expresia grafică a „distribuției normale") se referă la un anumit tip de distribuție teoretică care are câteva proprietăți caracteristice:

- are formă de „clopot". Cea mai mare parte a valorilor se concentrează în zona centrală (medie);
- este perfect simetrică pe ambele laturi ale sale;
- linia curbei se apropie la infinit de axa OX (orizontală), fără a o atinge vreodată;
- în conformitate cu proprietatea 2, de fiecare parte a mediei se află exact jumătate dintre valorile distribuției.

Exemple de curbe normale:

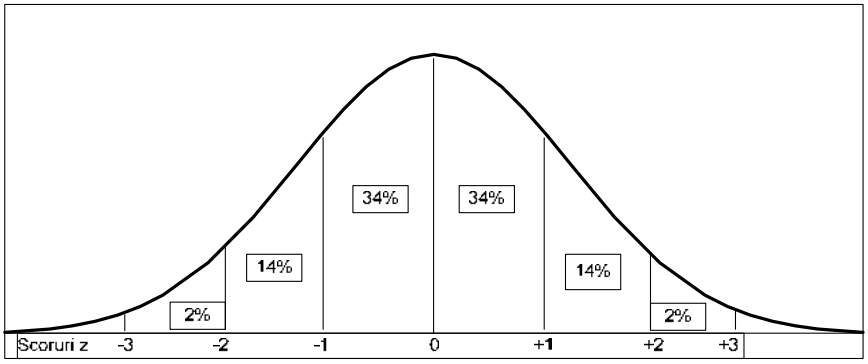




Imaginea de mai sus ilustrează diferite variante ale familiei de curbe normale, care respectă, fiecare dintre ele, condițiile de mai sus, chiar dacă au medii şi abateri standard diferite.

Distribuția normală z

Curba normală în care valorile sunt exprimate în scoruri *z* se numeşte **curba normală standardizată**. Ea are toate proprietățile enunțate mai sus, având însă şi parametrii oricărei distribuții *z*: m=0 şi s=1. Rezultă astfel că distribuția normală standardizată (*z*) este este simetrică în jurul lui 0.



Curba normală standardizată are câteva caracteristici care sunt figurate în imaginea de mai sus şi pe care este important să le reținem:

- Aproximativ 34% dintre scorurile distribuției normale se află între medie şi o abatere standard deasupra mediei (*z*=+1)
- Între – 1*z* şi +1*z* se află aproximativ 68% dintre valorile distribuției
- Aproximativ 96% dintre scoruri se află între –2*z* şi +2*z*

Având în vedere distribuția scorurilor *z* pe o curbă normală standardizată, aceasta poate fi utilizată pentru a afla răspuns la întrebări precum: *Care este procentajul de valori care se află sub/peste o anumită notă z; între anumite note z; ori între medie şi o notă z? Care este nota z corespunzătoare unui anumit procentaj de valori*? Pentru a răspunde la aceste întrebări, se utilizează o tabelă specială care conține, sub formă de probabilități, frecvențele valorilor de sub curba normală *z* (Anexa 1).

Aşa cum vom vedea mai departe, curba normală are o importanță aparte pentru analiza statistică. Aceasta, deoarece se acceptă faptul că variabilele statistice s-ar distribui mai ales sub aceasta formă dacă ar fi efectuate un număr mare (tinzând spre infinit) de măsurări.

Copyright © DEPARTAMENT ID 2008

Exemple:

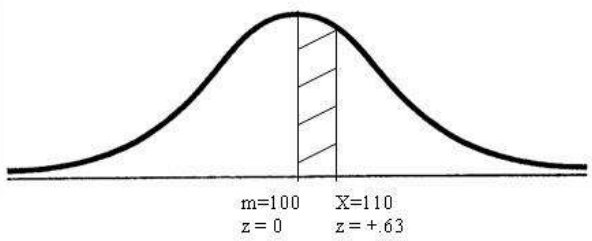
Să ne raportăm la distribuția valorilor QI, pentru care media este egală cu 100 şi abaterea standard 16

Exemplul 1: *Care este procentajul oamenilor al căror scor QI este între 100 şi 110?*

Pentru a răspunde la această întrebare, convertim valorile QI în scoruri *z*. 100(QI)=0(*z*). Pentru 110(QI) se aplică formula:

$$z = \frac{X - m}{s} = \frac{110 - 100}{16} = +0.63$$

Aria de sub curba normală cuprinsă între valorile QI şi 100 şi 110 este reprezentată pe figura următoare:



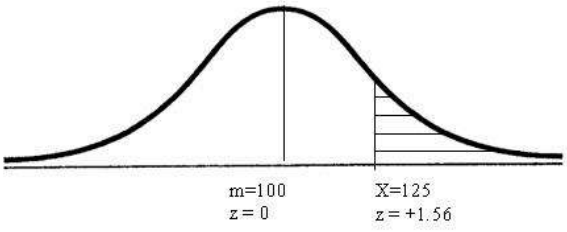
Citim tabela ariilor la intersecția celulelor 0.6 cu 0.03. Valoarea este 0.2357 ceea ce, exprimat în procente, este 23.57%

Conchidem că 23.57% din oameni au un QI cuprins între 100 şi 110)

Exemplul 2: *Care este procentul oamenilor al căror QI este mai mare decât 125?*
Convertim în note z:

$$z = \frac{X - m}{s} = \frac{125 - 100}{16} = +1.56$$

Aria de sub curba normală pentru scoruri QI mai mari decât 125 este reprezentată mai jos:



Citim valoarea din tabel care corespunde intersecției celulei 1.5 cu 0.06, pentru a afla procentajul dintre medie şi nota z +1.56. Găsim valoarea, exprimată în procente, 44.06%. Acesta este procentajul dintre medie şi z=+1.56.

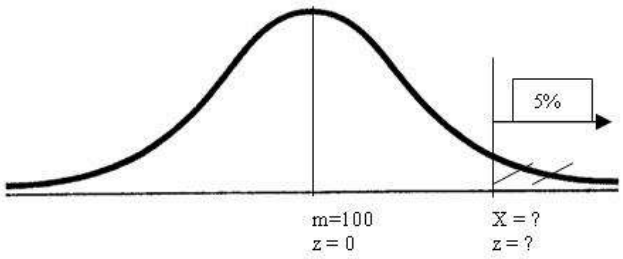
Ştim că procentajul peste medie este 50%, ca urmare, procentajul celor peste QI=125 va fi 50-44.06=5.94.

Conchidem că 5.94% dintre oameni au un QI mai mare de 125 (z=1.56)

Exemplul 3: *Care este scorul minim pe care trebuie să l obțină o persoană pentru a fi între primii 5% din populație?*

Ne reprezentăm aria de sub curbă care delimitează cele mai mari 5% dintre valorile z, trebuind să aflăm valoarea corespunzătoare z, respectiv QI:

Copyright © DEPARTAMENT ID 2008



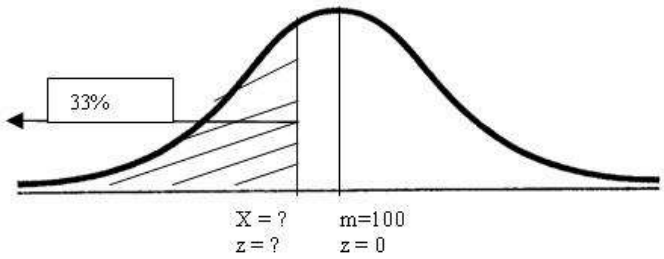
Aria dintre medie şi linia noastră este 50%-5%=45%. Căutăm în tabel valoarea cea mai apropiată de 0.45 şi o găsim la intersecția celulelor 1.6 cu 0.04. Deci, z=1.64 pentru limita procentului de 5%.

Convertim scorul z=1.64 în valoare brută: X=m+z*s=100+ (+1.64)*16=126.24

Conchidem că pentru a fi în primii 5% trebuie să obținem un QI=126.24

Exemplul 4: *Care este scorul care indică cei mai slabi 33%?*

Ne reprezentăm limita de 33% în zona valorilor de sub medie:



Căutăm scorul ***z*** corespunzătoare acestui procent.

Mai întâi, scădem 33% din 50% cât reprezintă aria din partea inferioară a curbei. Obținem 17% Căutăm nota z corespunzătoare procentului de 17% de sub medie.

Valoarea 0.1700 (17%) se găseşte la intersecția celulelor 0.4 cu 0.04, ceea ce indică nota *z*=-0.44 (cu minus, pentru că ne aflăm în partea stângă a curbei).

Convertim nota *z* în valoare brută: X=m+*z**s=100+(-0.44)*16=92.96.

Conchidem că este necesar un scor de cel mult 92.96 pentru a avea un QI între ultimii 33%.

Aria de sub curba normală văzută ca probabilitate

Valorile reprezentate pe curba normală nu reprezintă valori reale, rezultate în urma unui proces de măsurare. Ele reprezintă valori ipotetice, distribuite astfel pe baza unui model matematic (legea numerelor mari). Nimic nu ne împiedică să considerăm că valorile de sub curba normală sunt rezultatul unei ipotetice extrageri aleatoare. Pe măsură ce „extragem" mai multe valori, curba de distribuție a acestora ia o formă care se apropie de forma curbei normale. Extrăgând „la infinit" valori aleatoare, vom obține o distribuție normală perfectă, exprimabilă printr-o curbă normală perfectă.

Din cele spuse mai sus, rezultă faptul că valorile din zona centrală a curbei sunt mai „frecvente" (mai multe), pentru că apariția lor la o extragere aleatoare este mai „probabilă". În același timp, valorile „mai puțin probabile", apar mai rar şi populează zonele laterale, din ce în ce mai extreme, ale distribuției (curbei). Probabilitatea înseamnă „frecvența relativă a apariției unui eveniment". Subiectiv, se traduce prin „cât de siguri putem fi că acel eveniment apare". Dacă probabilitatea reprezintă raportul dintre evenimentul favorabil şi toate evenimentele posibile, atunci valoarea ei variază între 0 şi 1. Ea poate fi exprimată şi în procente. De exemplu, probabilitatea de 0.05 corespunde unui procentaj de apariție de 5%

Utilizând simbolul ***p*** *(de la „probabilitate")*, spunem că dacă p<0.05 înseamnă că evenimentul are mai puțin de 5% şanse să apară, în condițiile unei distribuții corespunzătoare curbei normale. Procentajul ariilor de sub curba normală poate fi citit deci, şi ca probabilitate a distribuției. De exemplu, probabilitatea de a avea un scor între medie şi ***z***=+1 este de p=0.34, iar probabilitatea de avea un scor ***z***=+1.65 sau mai mare, este mai mică de 0.05 (p<0.05).

Copyright © DEPARTAMENT ID 2008

Distribuții reale şi distribuții normale z

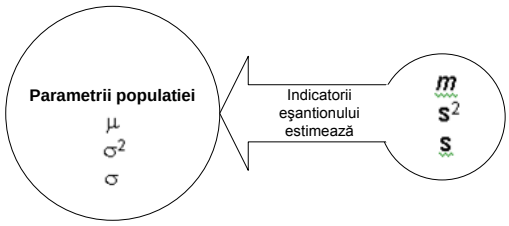
Caracteristicile curbei normale şi frecvența cu care se face apel la aceasta în studiile statistice determină adesea interpretări greşite. De aceea se cuvine să insistăm asupra faptului că distribuția normală reprezintă un model teoretic care se consideră că aproximează de o manieră mulțumitoare cele mai multe dintre distribuțiile caracteristicilor naturale, incluzându-le şi pe cele psihice. Cu toate acestea, distribuțiile reale pe care le descoperă psihologii în studiile lor nu au niciodată parametrii unei curbe normale perfecte. Acest lucru este practic imposibil dacă ne gândim că o curbă normală are limitele deschise, mergând spre infinit, în timp ce distribuțiile reale sunt întotdeauna finite. În ciuda acestui neajuns, aproximarea oferită de modelul teoretic al curbei normale este considerată acceptabilă din punct de vedere ştiințific.

Un alt aspect care poate conduce la interpretări eronate este exprimarea valorilor curbei normale în scoruri *z*. Acest fapt este înțeles adesea cu sensul că transformarea în scoruri *z* a unei distribuții o transformă automat într-o distribuție normală, ceea ce este o concluzie profund greşită. Convertirea valorilor unei distribuții în scoruri *z* nu modifică forma distribuției. Distribuția normală *z* este o distribuție teoretică, în timp ce o distribuție *z* oarecare are forma distribuției valorilor originale.

Distribuția de eşantionare

Populație şi eşantion

Obiectivul legitim al cercetării ştiințifice este identificarea unor adevăruri cu un anumit grad de generalitate. Din punct de vedere statistic „generalul" este reprezentat de totalitatea valorilor care descriu o anumită caracteristică, şi este numit „populație". Din păcate însă, investigarea tuturor „indivizilor" (valorilor) care compun o anumită populație nu este aproape niciodată posibilă. Ca urmare, în practica cercetării ştiințifice se supun cercetării psihologice loturi mai restrânse, numite eşantioane, extrase din ansamblul populației vizate. Parametrii descriptivi ai acestor eşantioane (medie, abatere stadard) sunt extrapolați, în anumite condiții şi cu ajutorul unor proceduri specializate, la populația din care fac parte.



A fundamenta un adevăr statistic înseamnă a trage o concluzie care descrie parametrii unei populații de valori, pe baza indicatorilor unui eşantion din acea populație.

În contextul cercetării statistice utilizăm următoarele definiții:

Populația reprezintă totalitatea „unităților de informație" care constituie obiectivul de interes al unei investigații. Prin „unități de informație" înțelegem cel mai adesea „persoane" („subiecți", cu un termen uzual in cercetarea psihologică[2]). Dar, la fel de bine, putem înțelege şi „populația de cupluri familiale", sau „populația" de diferențe dintre mediile a două variabile, de exemplu. În esență, prin „populație" trebuie să înțelegem extinderea maximă posibilă, sub aspectul volumului, a respectivei „unități de informație". Extinderea menționată este, la rândul ei, definită prin obiectivul de cercetare, ceea ce înseamnă ca are o dimensiune subiectivă. Aceasta se referă la domeniul de interes pe care şi-l propune cercetătorul. De exemplu, într-un studiu cu privire la efectul oboselii asupra performanței cognitive, pot fi vizate diferite categorii de „populații": a aviatorilor, a studenților, a mecanicilor de locomotivă, a şahiştilor, etc. Este de la

[2] Denumirea de „subiect" este respinsă în mediul anglo-saxon din cauza semnificației de „supus" pe care o are în limba engleză cuvântul „subject". Din acest motiv, o variantă frecvent uzitată este cea de „participant".

Copyright © DEPARTAMENT ID 2008

sine înțeles faptul că, încă de la începutul unei cercetări științifice, se va preciza populația cercetării, cu alte cuvinte, domeniul de extindere a rezultatelor şi a concluziilor ce urmează a fi trase.

Eşantionul reprezintă „unitățile de informație" selecționate pentru a fi efectiv studiate. Ideea pe care se bazează cercetările bazate pe eşantioane este aceea că se pot face aprecieri asupra unei întregi populații, în anumite condiții, doar pe baza caracteristicilor măsurate pe o parte a acesteia.

Exemple:

- Într-un studiu asupra efectelor accesului la internet asupra elevilor de liceu, elevii de liceu reprezintă „populația", iar elevii selecționați pentru investigație, „eşantionul".

- Într-un studiu care vizează influența inteligenței asupra performanței în instruirea de zbor, populația este reprezentată de toți piloții, iar eşantionul, de subiecții incluşi în studiu.

Dacă am reuşi recoltarea datelor cu privire la întreaga populație care face obiectul cercetării, am putea trage concluzii directe cu privire la aceasta prin utilizarea indicatorilor statistici descriptivi cunoscuți (medie, dispersie, abatere standard) numiți şi „parametrii populației". Dar acest lucru nu este aproape niciodată posibil şi, ca urmare, indicatorii statistici ai eşantionului sunt utilizați pentru a face estimări, inferențe, cu privire la parametrii populației. În esență, a testa o ipoteză statistică înseamnă a emite concluzii asupra unei „populații" pe baza rezultatelor obținute pe un eşantion care aparține acelei populații. În acest context, demersul științific presupune următorii paşi:

1. formularea problemei cercetării (sub forma unei întrebări, cu referire la o anumită populație);
2. emiterea unei ipoteze privind cel mai probabil răspuns;
3. selectarea unui eşantion;
4. aplicarea unei proceduri care sa permită acceptarea sau respingerea ipotezei.

Reprezentativitatea eşantionului

Verificarea statistică a ipotezelor se bazează pe o idee simplă: dacă avem un eşantion a cărui alegere respectă anumite condiții, extras dintr-o populație oricât de mare, rezultatele obținute pe acesta pot fi extrapolate la întreaga populație.

Calitatea unui eşantion de a permite extinderea concluziilor la întreaga populație din care a fost extras se numeşte reprezentativitate. De fapt, nici un eşantion nu poate reprezenta perfect datele populației. De aceea reprezentativitatea are o semnificație relativă. Ca urmare estimările pe bază de eşantion conțin întotdeauna o doză mai mare sau mai mică de eroare. Cu cât eroarea este mai mică, cu atât concluziile obținute pe eşantion pot fi generalizate mai sigur asupra populației.

Pentru a permite fundamentarea inferențelor statistice, eşantionul trebuie să fie constituit din „unități de informație" (subiecți, valori, etc.) independente unele de altele. Independența valorilor se referă la faptul că fiecare valoare (sau unitate experimentală) trebuie să fie absolut distinctă de celelalte. În esență constituirea unui eşantion trebuie să evite efectele unor factori sistematici care să interfereze cu obiectivele studiului, orientând rezultatele într-o anumită direcție (situație desemnată în limba engleză prin termenul de *bias*).

Câteva exemple:

- Dacă măsurăm timpul de reacție la un număr de cinci subiecți, dar facem trei evaluări la fiecare subiect, nu avem eşantion de 15 valori independente, deoarece valorile aceluiaşi subiect au în comun o „constantă personală" care le face dependente una de cealaltă. Pentru avea un singur eşantion am putea să utilizăm media celor trei determinări pentru fiecare subiect.
- Dacă dorim să investigăm efectul inteligenței asupra performanței şcolare trebuie să avem grijă să includem în eşantion subiecți provenind din familii cu un nivel variat al veniturilor, pentru a anihila influența statutului socio-economic asupra performanței şcolare.
- Un studiu asupra atitudinii față de utilizarea computerelor în educație, poate fi influențat în mod sistematic dacă eşantionul este constituit numai din elevi care utilizează frecvent calculatorul.
- În cazul unui sondaj cu privire la intențiile de vot bazat pe interviul telefonic, vom obține rezultate afectate de starea socială a respondenților (îşi permit montarea unui telefon) sau de ora apelului (în orele dimineții sunt acasă, să zicem, mai multe femei casnice).

Este clar de ce modul de constituire a eşantionului este decisiv pentru nivelul de reprezentativitate. Esențială în acest caz este asigurarea condițiilor ca acesta să acopere în mod real caracteristicile populației,

Copyright © DEPARTAMENT ID 2008

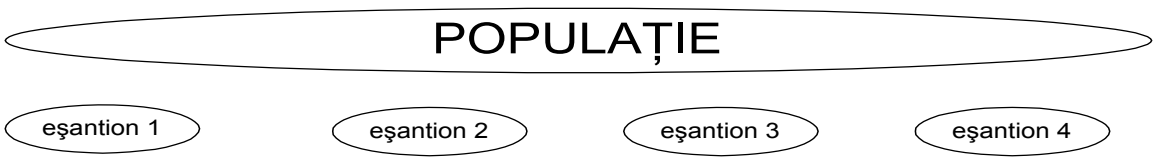
evitându-se „favorizarea" sistematică a unor subiecți „nereprezentativi". Fără a intra în amănunte tehnice cu privire la procedurile de eşantionare, iată care sunt cele mai utilizate metode de constituire a eşantioanelor:

a) Eşantionare stratificată multistadială. Populația se împarte în categorii, fiecare categorie în subcategorii ş.a.m.d., iar subiecții sunt selecționați aleator la nivelul categoriei de nivelul cel mai scăzut. Se obține astfel un eşantion care reproduce fidel structura populației.
b) Eşantionare prin clasificare unistadială. Se identifică categorii pe un singur nivel iar subiecții se extrag aleator din fiecare categorie.
c) Eşantionare aleatoare. Subiecții sunt extraşi aleator (la întâmplare) din ansamblul populației. „La întâmplare", înseamnă în acest caz utilizarea unei proceduri care asigură fiecărui subiect al populației absolut aceleaşi şanse de a fi extras. În acest scop se pot utiliza programe de calculator (de ex. SPSS) sau tabele de numere aleatoare.
d) Eşantionare pseudo-aleatoare (*haphazard,* sau de convenienţă). Sunt utilizați subiecții „disponibili". Este cazul cel mai frecvent întâlnit în practică şi, dacă „disponibilitatea" nu este afectată de un aspect care să influențeze semnificativ obiectivul cercetării, atunci reprezentativitatea este acceptabilă.

În concluzie, presupunând că am obținut anumite rezultate pe un eşantion aleator, raționamentul statistic ne permite să aplicăm concluziile la întreaga populație din care a fost extras acel eşantion. Se impune însă, o precizare clară a populației de referință pentru că, dincolo de limitele acesteia, extrapolarea nu este permisă. De exemplu, rezultatele unui studiu asupra atitudinii față de internet efectuat pe un eşantion de studenți nu poate fi extrapolat la alte categorii sociale, şi nici chiar la alte categorii de studenți, dacă în eşantionul nostru au intrat numai studenți de la facultăți umaniste, să zicem.

Distribuția mediei de eşantionare

Atunci când constituim un eşantion de studiu nu facem decât să utilizăm doar unul dintre eşantioanele posibil a fi selecționate (alese, constituite, extrase) din populația cercetării. Dacă am selecta mai multe eşantioane din aceeaşi populație, fiecare dintre ele ar fi caracterizat prin indicatori sintetici specifici, vor avea, fiecare, media şi abaterea lor standard. Imaginea de mai jos sugerează situația descrisă:



Dacă fiecare dintre cele patru eşantioane de valori are propria sa medie, atunci distribuția mediilor tuturor eşantioanelor extrase se numeşte *distribuția mediei de eşantionare* sau, mai scurt, *distribuția de eşantionare*. La rândul ei, distribuția mediilor are şi ea o medie, numită *medie de eşantionare*, şi care se calculează, evident, după următoarea formulă:

$$\mu = \frac{m_1 + m_2 + m_3 + ... + m_k}{k}$$

unde μ este media populației, valorile *m* sunt mediile fiecărui eşantion constituit, iar *k* este numărul eşantioanelor.

Dacă am extrage toate eşantioanele posibile dintr-o populație, atunci media de eşantionare este identică cu media populației. Pentru exemplificare, să presupunem că avem o „populație" constituită din valorile 1,2,3,4 şi să ne propunem constituirea tuturor eşantioanelor posibile de câte 3 valori. Tabelul de mai jos ilustrează această situație:

Populația	**Eşantioane**	**Distribuția mediei de eşantionare**
1	1,2,3	m_1=2.00
2	1,2,4	m_2=2.33
3	3,4,1	m_3=2.67

Copyright © DEPARTAMENT ID 2008

4	2,3,4	m_4=3.00
μ=2.5 **σ=1.29**	*Toate eşantioanele posibile pentru N=3*	**Σ=10.00** **m=10/4=2.5**

Aşa cum se observă, dacă extragem toate eşantioanele posibile (în acest caz 4) dintr-o populaţie de valori, atunci media mediilor eşantioanelor extrase (denumită *medie de eşantionare*) este identică cu media populaţiei (în cazul dat: m=μ=2.5). Datele din tabel ne mai arată şi faptul că media fiecărui eşantion oscilează (variază) în jurul mediei de eşantionare. De aceea ele pot fi considerate o estimare a acesteia din urmă, în ciuda impreciziei pe care o conţine fiecare. Această imprecizie se numeşte eroare de estimare. Desigur, exemplul are o valoare de ilustrare teoretică deoarece, în practică, niciodată nu se ajunge la selectarea tuturor eşantioanelor posibile dintr-o anumită populaţie de valori.

Împrăştierea distribuţiei de eşantionare (eroarea standard a mediei)

Distribuţia de eşantionare nu are aceeaşi împrăştiere ca şi distribuţia valorilor individuale ale variabilei de origine. Aceasta pentru că, la nivelul fiecărui eşantion, o parte din împrăştierea totală este „absorbită" de media fiecărui eşantion în parte. Cu cât eşantioanele sunt mai mari, cu atât media fiecărui eşantion tinde să fie mai apropiată de media variabilei originale şi, implicit, abaterea standard a distribuţiei de eşantionare este mai mică prin comparaţie cu abaterea standard a variabilei.

Exemplu: Să considerăm populaţia valorilor 1,2,3,4,5,6,7,8,9,10, pentru care am calculat μ=5.5 şi σ=3,0276. Am extras, cu ajutorul unui program statistic, cinci eşantioane aleatoare (pentru uşurinţa calculelor, am ales pentru fiecare eşantion N=3). Iată cum se prezintă mediile şi abaterile standard pentru cele cinci eşantioane selectate:

m_1=5.00	m_2=4.5	m_3=4.0	m_4=2.5	m_5=5.5
s_1=5.65	s_2=4.94	s_3=4.24	s_4=2.12	s_5=6.36

În acest exemplu, cele cinci eşantioane nu sunt toate, ci doar o parte din eşantioanele posibile de 3 valori extrase din populaţia cercetată. Media distribuţiei de eşantionare pentru acest exemplu este:

$$\mu = \frac{m_1 + m_2 + m_3 + m_4 + m_5}{5} = 5.375$$

În ceea ce priveşte împrăştierea distribuţiei de eşantionare, aceasta este, aşa cum am spus, mai mică decât împrăştierea variabilei la nivelul întregii populaţii, deoarece o parte a împrăştierii generale se concentrează (se „pierde") în media fiecărui eşantion extras. Ca urmare, abaterea standard a distribuţiei de eşantionare este o fracţiune din abaterea standard a populaţiei, fiind dependentă de mărimea eşantionului. Mai precis, fără a intra în detalii explicative, abaterea standard a distribuţiei de eşantionare este egală cu $\sqrt{N}$ din abaterea standard a populaţiei, unde N este volumul eşantionului.

Deoarece împrăştierea mediei de eşantionare arată cât de mult se abat aceste medii de la media populaţiei, abaterea standard a mediei de eşantionare este denumită *eroare standard a mediei* şi se calculează cu formula:

$$s_m = \frac{\sigma}{\sqrt{N}} \qquad \text{(formula 3.1)}$$

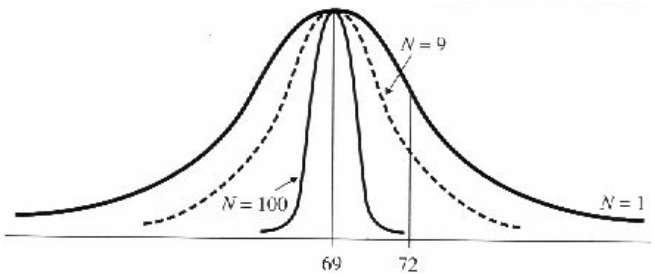
unde $\mathbf{s_m}$ este eroarea standard a mediei de eşantionare, σ este abaterea standard a populaţiei iar N este volumul eşantionului. În cazul distribuţiei de mai sus, eroarea standard a mediei este

$$s_m = \frac{\sigma}{\sqrt{N}} = \frac{3.02}{\sqrt{3}} = 1.74$$

Pentru că, în mod obişnuit, abaterea standard a populaţiei nu este cunoscută, eroarea standard a mediei de eşantionare se calculează utilizând abaterea standard a eşantionului, care reprezintă o estimare a împrăştierii la nivelul populaţiei.

Figura de mai jos sugerează foarte bine modul în care, prin creşterea volumului eşantionului, media eşantionului se apropie tot mai mult de media populaţiei, cu alte cuvinte, comportă o eroare din ce în ce în mai mică faţă de aceasta.

Copyright © DEPARTAMENT ID 2008



Expresia de „*eroare standard a mediei*" poate fi mai greu de înțeles, dat fiind faptul că este folosită pentru a defini un indicator al împrăştierii, în timp ce are în compunere cuvântul „medie". Trebuie însă să reținem faptul că acest indicator măsoară cât de departe poate fi media unui eşantion de media populației din care a fost extras. Altfel spus, câtă „eroare" poate conține media unui eşantion în estimarea mediei populației. Având în vederea faptul că la numitor avem o expresie bazată pe N (volumul eşantionului), este limpede de ce, cu cât eşantionul este mai mare, cu atât eroarea standard a mediei este mai mică.

Teorema limitei centrale[3]

Teorema limitei centrale certifică două adevăruri statistice fundamentale:

1. Cu cât numărul eşantioanelor realizate dintr-o populație (tinzând spre infinit) este mai mare, cu atât media distribuției de eşantionare se apropie de media populației.
2. Distribuția mediei de eşantionare se supune legilor curbei normale, chiar şi atunci când distribuția variabilei la nivelul întregii populații nu are un caracter normal, cu condiția ca volumul eşantioanelor să fie „suficient de mare". Cu alte cuvinte, distribuția mediei de eşantionare se apropie de distribuția normală, cu atât mai mult cu cât volumul eşantionului este mai mare.

   Teorema limitei centrale este adevărată în următoarele condiții fundamentale:
   a. eşantioanele sunt aleatoare sau neafectate de erori (bias);
   b. valorile care compun eşantioanele sunt independente unele de altele (măsurarea unei valori nu este influențată de măsurarea altei valori din eşantion);
   c. eşantioanele au acelaşi volum de valori (subiecți).

Utilitatea teoremei limitei centrale constă în faptul că ea permite fundamentarea inferențelor statistice fără a ne preocupa prea mult de forma distribuției valorilor individuale la nivelul populației. Este de ajuns să utilizăm un eşantion „suficient de mare" pentru a ne putea asuma presupunerea unei distribuții normale la nivelul mediei de eşantionare.

Întrebarea care se pune este, însă, cât de mare trebuie să fie un eşantion pentru a putea fi considerat „suficient de mare"? Fără a intra în amănunte, vom spune că, dacă eşantionul de referință cuprinde cel puțin 30 de subiecți, teoria statistică acceptă că avem o distribuție normală a mediei de eşantionare. Acest număr (30), care nu are nimic magic în el, este utilizat de obicei pentru constituirea eşantioanelor minime de cercetare. Pe această bază orice eşantion având cel puțin 30 de valori este considerat „eşantion mare" în timp ce orice eşantion cu mai puțin de 30 de valori este considerat „eşantion mic".

În concluzie, distribuția mediei de eşantionare are o evoluție diferită de distribuția valorilor individuale ale unei caracteristici. Chiar şi atunci când acestea din urmă nu se distribuie după regulile curbei normale, mediile eşantioanelor tind spre o distribuției normală dacă volumul lor este suficient de mare. Mărimea eşantionului trebuie să fie de cel puțin 30 de valori pentru a avea încredere că teorema limitei centrale se verifică. Dar chiar şi eşantioane de volum mai mic pot avea medii ce se plasează pe o distribuție normală, dacă provin din populații normale. Din păcate, forma distribuției la nivelul populației nu este aproape niciodată cunoscută. În acest caz singurul lucru pe care îl putem face este să utilizăm, ori de câte ori ne putem permite, „eşantioane mari", adică de cel puțin 30 de valori, şi chiar mai mari, dacă acest lucru este posibil. Cu toate acestea, aşa cum vom vedea mai departe, există soluții statistice şi pentru eşantioane mai mici de 30 de valori[4].

[3] Sau „teorema limită centrală".
[4] Dincolo de aceste considerente teoretice, mărimea eşantioanelor utilizate în studiile statistice psihologice face obiectul unor recomandări specifice pentru diferite situații practice de cercetare. Acestea vor fi prezentate mai târziu.

Copyright © DEPARTAMENT ID 2008

Sarcina de lucru nr. 2. 1

Presupunem că evaluarea preferinței pentru risc la un grup de 30 de studenți aviatori care au suferit incidente critice în zbor a condus la o distribuție de valori având **m=60** şi **s=25**. Ştiind că indicele preferinței pentru risc la toți elevii piloți („populația") este **55**, şi are o distribuție normală, calculați răspunsul la următoarele întrebări:

1. Care este scorul z corespunzător eşantionului?
2. Care este procentajul valorilor între media e□antionului (60) şi media popula□iei?
3. Care este procentajul valorilor mai mari decât 60, raportat la media popula□iei?
4. Care este procentajul scorurilor mai mici de 60, raportat la media popula□iei?
5. Care este probabilitatea de avea un scor mai mare de 53, raportat la media popula□iei?
6. Care este probabilitatea de a avea un scor mai mic de 40, raportat la media popula□iei?
7. Care este probabilitatea de a avea un scor cuprins între 45 şi 48?
8. Care este scorul minim pe care îl poate avea o persoană pentru a intra în primii 10% dintre subiecți?
9. Care este scorul maxim pe care trebuie să îl obțină cineva pentru a se afla printre ultimii 15%?

Verificați răspunsurile corecte

Notele standardizate z pentru eşantioane (grupuri)

Ne vom referi acum la exemplul anterior, în care avem cinci eşantioane extrase dintr-o populație de 10 valori. Dacă avem media distribuției de eşantionare şi abaterea standard a acesteia (calculată ca eroare standard a mediei, cu formula 3.1), atunci putem exprima media unui eşantion oarecare, ca scor standardizat ***z***, într-o manieră similară cu scorul standardizat ***z*** pentru o valoare oarecare. Rostul acestei transformări ar fi acela de a vedea în ce măsură media eşantionului de studiu se îndepărtează de media populației de referință. Cu alte cuvinte, în ce măsură rezultatul obținut pe eşantion este unul „obişnuit" (mai aproape de media populației) sau unul „neobişnuit" (mai îndepărtat de media populației).

Formula de calcul este foarte asemănătoare cu formula lui *z* pentru valori individuale:

$$z = \frac{m - \mu}{s_m} \qquad \text{(formula 3.2)}$$

unde **m** este media eşantionului, **μ** media populației, iar $\mathbf{s_m}$ este eroarea standard a mediei.

Dacă presupunem că obiectul studiului îl face eşantionul 1, atunci putem calcula mai întâi eroarea standard a mediei, astfel:

$$s_m = \frac{\sigma}{\sqrt{N}} = \frac{3.02}{\sqrt{3}} = 1.74$$

În exemplul nostru, limitat la o populație cunoscută, am putut calcula abaterea standard a populației (σ=3.02), dar pentru situații reale, cu populații nelimitate, acest lucru nu este posibil. În astfel de cazuri se acceptă faptul că abaterea standard a populației este „suficient de bine reprezentată" de abaterea standard a eşantionului extras din aceasta. Ca urmare, dacă nu aveam abaterea standard a populației, am fi putut utiliza

Copyright © DEPARTAMENT ID 2008

în formula erorii standard a mediei abaterea standard a eşantionului (în cazul nostru $s_{1=}5.65$ în loc de $\sigma=3.02$).
Mai departe, scorul standard *z* pentru eşantionul 1, se calculează astfel:

$$z = \frac{m - \mu}{s_m} = \frac{5 - 5.5}{1.74} = \frac{-0.5}{1.74} = -0.28$$

unde **m** este media eşantionului 1, **μ** este media populației, iar $\mathbf{s_m}$ este eroarea standard a mediei.

Exemplu:

Să presupunem că, la un examen de cunoştințe de statistică, o grupă de 45 de studenți obține un scor mediu de m=28.5 puncte. Presupunând că media pe populația studențească care a mai dat acest examen (calculată de-a lungul anilor anteriori) este μ=27.3, cu o abatere standard σ=8.2, trebuie să aflăm care este performanța grupei respective transformată în notă *z*.
Calculăm mai întâi abaterea standard a mediei:

$$s_m = \frac{\sigma}{\sqrt{N}} = \frac{8.2}{\sqrt{45}} = \frac{8.2}{6.70} = 1.22$$

Calculăm apoi scorul *z* pentru grup:

$$z = \frac{m - \mu}{s_m} = \frac{28.5 - 27.3}{1.22} = \frac{1.20}{1.22} = 0.98$$

Dacă vrem să ştim unde se plasează performanța grupului nostru pe o curbă normală, atunci ne uităm pe tabela notelor **z** şi găsim, în dreptul scorului z=0.98, valoarea tabelară 0.3365. Aceasta poate fi interpretat în mai multe feluri. De exemplu, putem spune că procentul performanțelor posibile peste nivelul grupului nostru este 50%-33%, adică 17%. Sau, în termeni probabilistici, putem sune şi că: „*probabilitatea de a avea o grupă (un eşantion, de aceeaşi mărime) care să obțină un scor mai bun la un examen de statistică (cu aceleaşi întrebări) este de 0.17"*.

Ipoteze şi decizii statistice

Să ne imaginăm că un psiholog şcolar îşi pune întrebarea dacă elevii participanți la olimpiadele şcolare au un nivel de inteligență (QI) superior elevilor în general. Dacă acceptăm că această problema prezintă interes din punct din vedere practic-pedagogic sau ştiințific, atunci se justifică transformarea ei într-o problemă de cercetare. În esență, această problemă ar putea fi formulată astfel: „*Elevii participanți la olimpiade sunt mai inteligenți decât toți elevii în general, fie ei participanți sau nu la olimpiade?*".

Ipoteza cercetării

În mod obişnuit, o cercetare ştiințifică se bazează pe estimarea unui rezultat aşteptat, denumit ipoteză. În cazul nostru, psihologul se poate aştepta în mod legitim ca participanții la olimpiadă să fie mai inteligenți decât elevii în general. Acest rezultat „aşteptat", „prefigurat", se numeşte ipoteza cercetării, fiind codificată cu H_1. Am putea formaliza ipoteza cercetării astfel:

$$H_1 \rightarrow m_{po} \neq m_{eg}$$

unde m_{po} reprezintă media inteligenței populației participanților la olimpiade, iar m_{eg} reprezintă media inteligenței populației elevilor în general.

În conformitate cu ipoteza cercetării, există două populații distincte sub aspectul nivelului de inteligență, cea a elevilor participanți la olimpiade şi cea a elevilor în general.

Copyright © DEPARTAMENT ID 2008

Ipoteza statistică (de nul)

Având în vedere că este imposibil să evalueze inteligența tuturor participanților la olimpiade, psihologul cercetător trebuie să găsească un răspuns la problema cercetării sale cu ajutorul unui eşantion. În acest scop, selectează la întâmplare, din populația de participanți la olimpiade, un grup de 30 de elevi, cărora le aplică un test de inteligență generală. Să presupunem că analiza rezultatelor indică pentru acest grup o medie a coeficientului de inteligență m=106 şi o abatere standard s=7. Amintindu-ne că media valorilor QI la nivelul întregii populații este μ=100 (σ=15)[5], se poate trage concluzia că elevii din populația de olimpici sunt mai inteligenți decât cei din populația generală de elevi? Aparent diferența de 6 unități QI în favoarea eşantionului cercetării i-ar îngădui o astfel de concluzie. Rigoarea ştiințifică îl obligă însă să observe că generalizarea mediei eşantionului de cercetare asupra întregii populații de elevi olimpici comportă anumite riscuri. Eşantionul cercetării, compus aleatoriu din elevi participanți la olimpiade, nu este decât unul din eşantioanele de olimpici care ar fi putut fi selectat. Astfel, faptul că eşantionul său are un QI mediu mai mare decât media populației se poate încadra în caracteristica oricărei medii de eşantion de a oscila în jurul mediei populației din care este extras. Ar fi posibil deci, ca valoarea medie de 106 să fie doar rezultatul hazardului, care face ca mediile eşantioanelor extrase din aceeaşi populație să varieze în jurul mediei populației.

Ca urmare, pentru a decide cu privire la ipoteza cercetării („*olimpicii sunt mai inteligenți decât elevii în general*") cercetătorul trebuie să evalueze probabilitatea ca media eşantionului cercetării să fie rezultatul hazardului de eşantionare. Rezultă de aici că, pentru a putea afirma că olimpicii sunt mai inteligenți decât media populației, cercetătorul trebuie să dovedească faptul că nivelul de inteligență al eşantionului de olimpici este mai mare decât al unui eşantion care ar fi fost extras absolut la întâmplare din populația generală de elevi.

Procedura statistică care se bazează pe acest raționament se numeşte „*ipoteză de nul*" (se utilizează şi alte variante: „*ipoteza diferenței nule*" sau, pur si simplu, „*ipoteză statistică*"). Respingerea ei implică o dovadă indirectă a validității ipotezei cercetării, şi se bazează pe un scenariu „negativ" (similar cu „a pune răul în față"). Ipoteza de nul se formulează ca opusul ipotezei cercetării. În cazul nostru ipoteza de nul va fi exprimată astfel: „*participanții la olimpiadă nu au o inteligență mai mare decât populația de elevi în general*".

Ipoteza de nul este simbolizată cu H_0, iar expresia ei formală este:

$$H_0 \rightarrow m_{po}=m_{eg}$$

ceea ce semnifică faptul că mediile celor două populații comparate nu diferă, ci sunt egale. Cu alte cuvinte, ipoteza de nul afirmă că nu există două populații distincte sub aspectul nivelului de inteligență, ci una singură. Elevii participanți la olimpiade nu se deosebesc sub aspectul inteligenței de populația elevilor în general.

Distribuția ipotezei de nul

Expresia $m_{po}=m_{eg}$ descrie situația în care media olimpicilor nu diferă de media populației generale de elevi, care poate fi definită, din acest motiv, drept „populația diferenței nule" sau, mai scurt, „populația de nul". Corespunzător, distribuția mediilor eşantioanelor aleatore extrase din populația de nul se numeşte „distribuția populației de nul" sau „distribuția de nul".

Aşa cum am spus anterior, extragerea unui număr mare de eşantioane (eventual infinit de mare), produce ceea ce se numeşte distribuția de eşantionare, care respectă legea curbei normale. Din perspectiva cercetării statistice, aceasta este chiar distribuția de nul, deoarece ilustrează forma în care se distribuie mediile tuturor eşantioanelor posibile, dacă acestea ar fi constituite pe o bază pur întâmplătoare, cu alte cuvinte, exact situația în care ipoteza de nul ar fi adevărată.

Dacă avem în vedere eşantioane extrase la întâmplare din populația de nul, atunci, în conformitate cu teorema limitei centrale, mediile acestora se distribuie pe o curbă normală. Ca urmare, putem utiliza tabela distribuției normale standard pentru a răspunde întrebărilor cu privire la media eşantionului de cercetare, în acelaşi mod în care am făcut-o pentru notele **z** individuale.

Dacă vrem să ştim care este probabilitatea de a obține un rezultat mai bun prin jocul şansei, nu trebuie decât să vedem unde se plasează rezultatul cercetării pe distribuția de nul. Apoi calculăm aria de dincolo de acest punct, deoarece aceasta ne arată proporția (probabilitatea) cazurilor în care eşantioane de

[5] În realitate, media QI este diferită în funcție de vârstă, dar, pentru exemplul nostru, vom accepta că populația generală de elevi are o medie de 100 şi o abatere standard de 15.

Copyright © DEPARTAMENT ID 2008

aceeaşi mărime, selectate la întâmplare din populația de nul, ar putea avea un QI mediu mai mare decât eşantionul de participanți la olimpiadă.

Testul *z* pentru un singur eşantion

Procedura de calcul

În urma aplicării testului de inteligență pentru eşantionul de participanți la olimpiadă (N=30) am obținut următoarele valori statistice: m=106 şi s=7. Ne amintim că media inteligenței populației, exprimată în unități QI, este μ=100, iar abaterea standard σ=15. Cu aceste date putem calcula nota *z* corespunzătoare eşantionului cercetării, cu formula:

$$z = \frac{m - \mu}{s_m} \quad \text{(formula 3.4)}$$

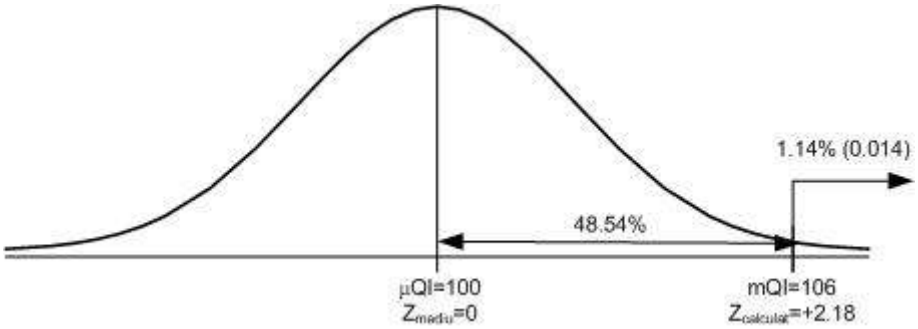
unde *m* este media eşantionului, μ este media populației, iar s_m este eroarea standard a mediei.

Rezultatul calculului este:

$$z = \frac{m - \mu}{s_m} = \frac{106 - 100}{\sigma / \sqrt{N}} = \frac{6}{15 / 5.47} = \frac{6}{2.74} = +2.18$$

În exemplul de mai sus, fiind vorba de o valoare QI, a cărei abatere standard la nivelul populației ne este cunoscută (am optat pentru σ=15) şi am utilizat-o ca atare. Dacă ar fi fost vorba de o variabilă pentru care nu cunoşteam abaterea standard la nivelul populației, am fi putut utiliza aceeaşi valoare calculată pe eşantionul de studiu (s=7).

Dacă citim frecvența corespunzătoare valorii *z* calculate (2.18) în tabelul distribuției normale, constatăm că între media populației de nul (z=0) şi nivelul inteligenței eşantionului de elevi olimpici se află 48.54% dintre valorile posibile. De aici rezultă că există 50-48.54 adică 1.46% şanse (sau o probabilitate p=0.0146) ca hazardul să producă un eşantion cu un QI egal sau mai mare decât eşantionul cercetării noastre. Imaginea de mai jos ilustrează grafic poziția mediei eşantionului de cercetare pe distribuția de nul.



Ne putem imagina o situație în care scorul mediu QI al eşantionului de participanți la olimpiadă este atât de mare încât să nu existe nici o şansă de a se obține un rezultat mai bun ca urmare a unei selecții întâmplătoare din populația de nul? Teoretic, acest lucru nu este posibil. Oricât de mare ar fi media unui eşantion de olimpici, hazardul poate produce un eşantion cu medie mai mare din populația de nul, deoarece curba normală este asimptotică. Există însă un „prag" dincolo de care probabilitatea unui eşantion aleatoriu din populația generală de elevi cu un QI mai mare decât cel al eşantionului de olimpici este atât de mică, încât să ne putem permite să o considerăm neglijabilă. Într-un asemenea caz, putem concluziona că valoarea calculată pe eşantionul cercetării nu decurge din variația întâmplătoare a mediei de eşantionare, ci provine din acțiunea unui factor sistematic care a condus la îndepărtarea semnificativă a mediei eşantionului de

Copyright © DEPARTAMENT ID 2008

studiu de media populației (în cazul nostru, accesul celor mai inteligenți elevi la olimpiadele şcolare). Despre „pragul" evocat mai sus, vom vorbi în continuare.

Decizia statistică

Următorul pas pe care trebuie să îl facă cercetătorul este acela de a decide dacă valoarea medie a eşantionului de olimpici decurge din faptul că aceştia sunt într-adevăr mai inteligenți decât elevii în general, sau reprezintă rezultatul unui joc al şansei, care a condus la selecția unui eşantion ce nu se diferențiază în mod real de populația de nul.

Este evident faptul că, dacă media eşantionului de olimpici ar fi fost egală cu 100, cercetătorul ar fi decis că valoarea nu confirmă ipoteza cercetării. În exemplul dat însă, media eşantionului cercetării fiind mai mare, ne punem problema, cât de mare trebuie să fie diferența față de media populației pentru a accepta că este o diferență „reală" (determinată de un factor de influență, accesul la olimpiadă pe baza inteligenței). Altfel spus, trebuie să decidem dacă acceptăm sau respingem ipoteza de nul.

Din păcate, nu există un criteriu obiectiv de decizie într-o situație de acest gen. Acceptarea sau respingerea ipotezei de nul depinde de gradul de risc pe care suntem dispuşi să ni-l asumăm în acest sens. Este evident că cineva interesat în acceptarea ideii că olimpicii sunt mai inteligenți ar fi dispus să considere că valoarea obținută este suficient de îndepărtată de medie pentru a respinge ipoteza de nul. La fel cum, cineva neîncrezător în această ipoteză (considerând că efortul de studiu, motivația, fac diferența dintre participanții şi neparticipanții la olimpiadele şcolare), ar putea fi dispus să impună un prag de respingere mult mai sever. Iată de ce, în practica cercetării ştiințifice s-a impus convenția unui prag maxim de risc acceptat pentru decizia statistică. Acest prag „critic" se numeşte *nivel alfa* (α) şi corespunde probabilității de 0.05. Pe curba normală *z*, fiecărei probabilități îi corespunde o anumită valoare *z*, ca urmare şi probabilității „critice" alfa îi corespunde o valoare *critică z*. Dat fiind faptul că a început prin a fi citită dintr-un tabel, mai este desemnată şi ca „valoare tabelară".

Avem acum toate elementele pentru luarea deciziei statistice în cazul cercetării noastre, pe baza unui raționament convențional, identic pentru întreaga comunitate ştiințifică. Esența acestuia constă în comparația rezultatelor derivate dintr-un context de cercetare cu cele specifice unui context ipotetic, aleatoriu (bazat pe şansa pură), după cum urmează:

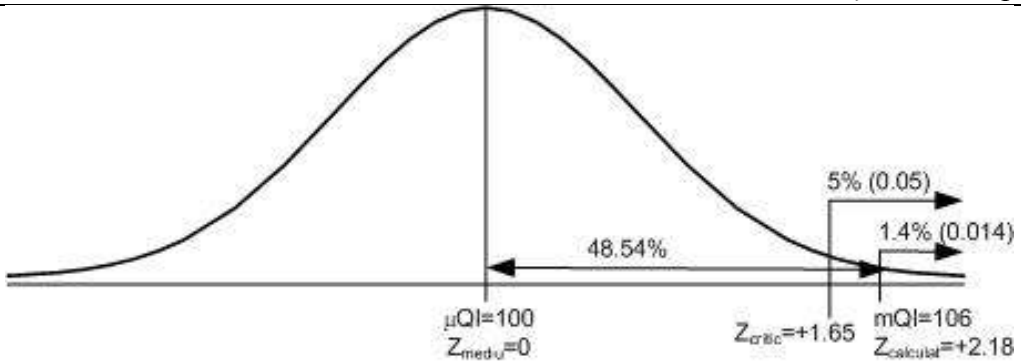
a. Dacă rezultatul calculat pentru eşantion este cel puțin egal sau mai mare decât scorul critic, atunci avem un rezultat semnificativ al cercetării. Aceasta, deoarece se acceptă că şansele ca acest rezultat să fi decurs din întâmplare sunt suficient de mici pentru a fi ignorate. În consecință, într-un astfel de caz, ipoteza de nul (H_0) se respinge, iar ipoteza cercetării (H_1) se consideră confirmată la un prag alfa=0.05 (dacă acesta a fost nivelul ales).

b. Dacă rezultatul eşantionului este mai mic decât scorul *z* critic, atunci avem un rezultat nesemnificativ al cercetării, prin faptul că există prea multe şanse ca acesta să poată fi obținut în condiții pur aleatoare. În această variantă, ipoteza de nul se acceptă, iar ipoteza cercetării se consideră infirmată la un prag alfa=0.05.

c. Cele două reguli decizionale de la punctele a şi b sunt exprimate pe baza comparației dintre valoarea calculată a testului şi valoarea critică tabelară, aferentă nivelului alfa. Ele însă pot fi exprimate şi direct, prin comparația probabilității valorii calculate cu alfa. Singura diferență este dată de faptul că raportul dintre probabilitatea asociată scorului calculat şi alfa este invers decât în cazul valorilor. Astfel, ipoteza de nul se admite dacă probabilitatea (p) a valorii calculate este mai mare decât alfa, şi se respinge dacă este egală sau mai mare decât acesta. Această precizare, îşi dovedeşte utilitatea în momentul în care se utilizează programe statistice, care fac inutilă consultarea tabelelor distribuției de nul, deoarece dau direct probabilitatea asociată valorii calculate a testului.

Imaginea de mai jos ilustrează poziția valorii calculate a testului *z* în raport cu valoarea critică pentru alfa=0.05.

Copyright © DEPARTAMENT ID 2008

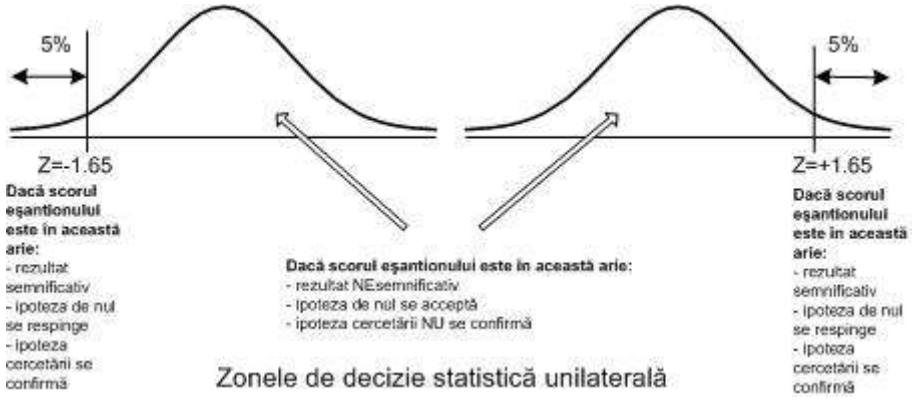


Dat fiind faptul că *z* calculat (+2.18) este mai mare decât *z* critic pentru valoarea lui alfa=0.05 (+1.65), decidem respingerea ipotezei de nul[6]. Ca urmare, în legătură cu studiul nostru demonstrativ, trebuie să decidem respingerea ipotezei de nul („*participanții la olimpiade nu sunt mai inteligenți decât elevii în general"*) ceea ce înseamnă, implicit, confirmarea ipotezei de cercetare. („*participanții la olimpiade sunt mai inteligenți decât elevii în general"*).

Raționamentul deciziei statistice exemplificat astfel, se va regăsi în toate situațiile de testare a ipotezelor statistice cu care ne vom confrunta mai departe, indiferent de modelul de cercetare şi de natura relației pe care vrem să o demonstrăm între variabile.

Decizii statistice unilaterale şi bilaterale

În exemplul nostru, ipoteza cercetării a fost aceea că elevii participanți la olimpiade au o inteligență mai mare decât media populației de nul. Din acest motiv, ne-a interesat să vedem în ce măsură rezultatul nostru confirmă ipoteza pe direcția valorilor din dreapta curbei normale (valori mari, cu *z* pozitiv). Ca urmare, am efectuat ceea ce se numeşte un test unilateral (*one-tailed*). În acest caz, ipoteza că participanții la olimpiadele şcolare ar putea avea o inteligență sub medie, nu este viabilă, dar dacă am fi obținut un z negativ pentru eşantionul cercetării, ar fi trebuit să îl testăm în partea din stânga curbei de distribuție, În aceste două situații am fi avut același *z* critic (1.65) cu semnul + sau – în funcție de zona scalei pentru care făceam testarea. Imaginea de mai jos ilustrează grafic cele două direcții de testare a ipotezelor statistice unilaterale şi ariile valorilor semnificative/nesemnificative, în funcție de valoarea critică a lui *z*.


5%
Z=-1.65
Dacă scorul eşantionului este în această arie:
- rezultat semnificativ
- ipoteza de nul se respinge
- ipoteza cercetării se confirmă
Dacă scorul eşantionului este în această arie:
- rezultat NEsemnificativ
- ipoteza de nul se acceptă
- ipoteza cercetării NU se confirmă
5%
Z=+1.65
Dacă scorul eşantionului este în această arie:
- rezultat semnificativ
- ipoteza de nul se respinge
- ipoteza cercetării se confirmă
Zonele de decizie statistică unilaterală

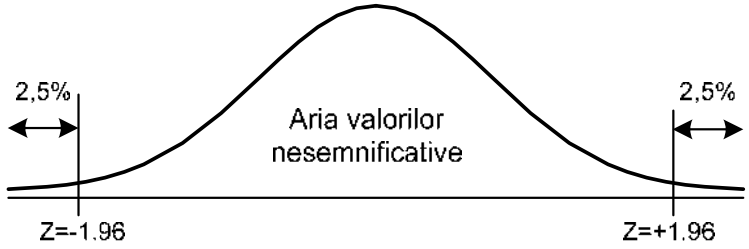
Ce s-ar fi întâmplat însă dacă eşantionul cercetării ar fi obținut un scor QI=94, ceea ce ar fi corespuns unui scor z=-2.18? În acest caz, aplicând un test unilateral orientat spre valori superioare mediei, conform ipotezei, ar fi trebuit să acceptăm ipoteza de nul, concluzionând că olimpicii nu sunt mai inteligenți decât media, fără a putea emite o concluzie privitoare la faptul că ei sunt, de fapt, mai puțin inteligenți, aşa cum ar fi cerut-o datele cercetării.

Pentru a elimina acest neajuns putem verifica ipoteza pe ambele laturi ale distribuției, aplicând ceea ce se numeşte un test bilateral (*two-tailed*). În acest caz se păstrează același nivel *alfa* (0.05), dar el se distribuie în mod egal pe ambele extreme ale curbei, astfel încât pentru 2.5% de fiecare parte, avem un **z**

[6] Puteam ajunge la aceeaşi concluzie pe baza faptului că probabilitatea valorii calculate (0.014) este mai mică decât alfa (0.05), dar acest raționament nu este posibil decât atunci când utilizăm programe specializate de calcul, care ne oferă direct valoarea lui p calculat.

Copyright © DEPARTAMENT ID 2008

critic de 1.96 (cu semnul - sau +). Această valoare este luată din tabelul ariei de sub curbă, în dreptul probabilității 0.4750 care corespunde unei probabilități complementare de 0.025 (echivalent cu 2.5%).



Zonele de decizie statistică bilaterală (p=0,05)

Figura de mai sus indică scorurile critice pentru un test **z** bilateral. Se observă că în cazul alegerii unui test bilateral (z=±1.96) nivelul α de 5% se împarte în mod egal între cele două laturi ale curbei. Este de la sine înțeles faptul că semnificația statistică este mai greu de atins în cazul unui test bilateral decât în cazul unui test unilateral, deoarece valoarea testului trebuie să fie mai mare de 1.65, cât este în cazul pentru un test unilateral.

Alegerea tipului de test, unilateral sau bilateral, este la latitudinea cercetătorului. De regulă însă, se preferă testul bilateral, chiar şi în situații de cercetare cum este aceea din exemplul nostru, când o diferență negativă față de media populației este improbabilă. Motivul îl constituie necesitatea de a introduce mai multă rigoare şi de a lăsa mai puțin loc hazardului. Se alege testul unilateral doar atunci când suntem interesați de evaluarea semnificației strict într-o anumită direcție a curbei, sau atunci când miza rezultatului este prea mare încât să fie justificată asumarea unui risc sporit de eroare. În mod uzual, ipotezele statistice sunt testate bilateral, chiar dacă ipoteza cercetării este formulată în termeni unilaterali. Testarea unilaterală este utilizată numai în mod excepțional, în cazuri bine justificate.

O scurtă discuție pe tema *nivelului alfa* maxim acceptabil (0.05) se impune, având în vedere faptul că întregul eşafodaj al deciziei statistice se sprijină pe acest prag. Vom sublinia, din nou, că p=0.05 este un prag de semnificație convențional, impus prin consensul cercetătorilor din toate domeniile, nu doar în psihologie. Faptul că scorul critic pentru atingerea pragului de semnificație este ±1.96 a jucat, de asemenea, un rol în impunerea acestei convenții. Practic, putem considera că orice îndepărtare mai mare de două abateri standard de la media populației de referință este semnificativă. Chiar dacă persistă posibilități de a ne înşela, ele sunt suficient de mici pentru a le trece cu vederea.

Impunerea unui prag minim de semnificație a testelor statistice are însă, mai ales, rolul de a garanta faptul că orice concluzie bazată pe date statistice răspunde aceluiaşi criteriu de exigență, nefiind influențată de subiectivitatea cercetătorului. Nivelul alfa de 0.05 nu este decât pragul maxim acceptat. Nimic nu împiedică un cercetător să îşi impună un nivel mai exigent pentru testarea ipotezei de nul, ceea e înseamnă un prag alfa mai scăzut. În practică mai este utilizat pragul de 0.01 şi, mai rar, cel de 0.001. Toate aceste praguri pot fi exprimate şi în procente, prin opusul lor, care exprimă *nivelul de încredere* în rezultatul cercetării. Astfel, printr-o probabilitate de 0.05 se poate înțelege şi un nivel de încredere de 95% în rezultatul cercetării (99%, pentru p=0.01 şi, respectiv, 99.9% pentru p=0.001).

În fine, este bine să subliniem faptul că utilizarea acestor „praguri" vine din perioada în care nu existau calculatoare şi programe automate de prelucrare statistică. Din acest motiv, cercetătorii calculau valoarea testului statistic pe care apoi o comparau cu valori tabelare ale probabilității de sub curba de referință. Pentru a face mai practice aceste tabele, ele nu cuprindeau toate valorile de sub curbă, ci doar o parte dintre acestea, printre ele, desigur, cele care marcau anumite „praguri". Rezultatul cercetării era raportat, de aceea, prin invocarea faptului de a fi „sub" pragul de semnificație sau „deasupra" sa. Odată cu diseminarea pe scară largă a tehnicii de calcul şi cu apariția programelor de prelucrări statistice, semnificația valorilor testelor statistice nu mai este căutată în tabele, ci este calculată direct şi exact de către program, putând fi afişată ca atare. De aici, aşa cum am mai spus, rezultă şi posibilitatea de a lua decizia statistică prin compararea directă a valorii calculate a lui *p* cu pragul alfa critic asumat.

Estimarea intervalului de încredere pentru media populației

Eşantionul cercetării noastre a obținut medie QI=106, care s-a dovedit semnificativă. Acest lucru înseamnă că valorile inteligenței elevilor olimpici fac parte dintr-o populație specială de valori QI, care are o medie mai mare decât media populației generale de elevi. Dar cât de mare este această medie? Media

Copyright © DEPARTAMENT ID 2008

eşantionului cercetării ne oferă o estimare a acesteia dar, ca orice estimare, conține o anumită imprecizie, exprimată prin eroarea standard a mediei. Nu vom putea şti niciodată cu precizie care este media inteligenței populației de elevi olimpici, dar teorema limitei centrale ne permite să calculăm, cu o anumită probabilitate, în ce interval se află ea, pe baza mediei eşantionului cercetării şi a erorii standard a acesteia.

Acest lucru se bazează pe proprietatea curbei normale de a avea un număr bine definit de valori pe un interval simetric în jurul mediei. Astfel, dacă luăm pe curba normală un interval cuprins între z=±1.96 de o parte şi de alta a mediei, ştim că acoperim aproximativ 95% din valorile posibile ale distribuției. În acest caz, z=±1.96 se numeşte ***z* critic** deoarece reprezintă un prag limită, pe cele două laturi ale distribuției (care, pentru curba normală standardizată, este 0). Alegerea acestor limite pentru ***z* critic** este convențională. Se pot alege, la fel de bine, valori simetrice ale lui z care să cuprindă între ele 99% sau 99.9% dintre valorile de pe curba normală. Prin consens, însă, se consideră că asumarea unui nivel de încredere de 95% (corespunzător pentru valori „critice" ale lui z=±1.96) este considerat suficient pentru păstrarea unui echilibru între precizia estimării şi probabilitatea estimării. Ca urmare, în această condiție, putem spune că există 95% şanse ca, având media unui eşantion aleator, media populației să se afle undeva în intervalul:

$$\mu = m \pm z_{critic} * s_m \qquad \text{(formula 3.3)}$$

unde μ=media populației, pe care o căutăm
m=media eşantionului de cercetare
z_{critic}=valoarea corespunzătoare pentru alfa ales (de regulă 0.05)
s_m=eroarea standard a mediei

În ce priveşte eroarea standard a mediei, aceasta este dată de raportul dintre abaterea standard a populației, pe care în acest caz o cunoaştem (15) şi radical din volumul eşantionului:

$$s_m = \frac{15}{\sqrt{30}} = 2.74$$

Mai departe, utilizând formula 3.3 pentru datele eşantionului cercetării, limitele de încredere pentru media populației mediei pot fi calculate astfel:

pentru limita inferioară $\mu = 106 - 1.96 * 2.74 = 100.62$
pentru limita superioară $\mu = 106 + 1.96 * 2.74 = 111.37$

Ca urmare, putem afirma, cu o probabilitate de 95%, că media reală a populației de elevi olimpici, estimată prin media eşantionului cercetării, se află undeva între 100.6 şi 111.3. Acest interval a cărui limită inferioară este foarte aproape de media populației generale de valori QI (100), ne arată că, deşi semnificativă, diferența eşantionului nostru nu are o valoare foarte ridicată. Trebuie să observăm, de asemenea, că mărimea intervalului de încredere rezultă din imprecizia mediei, exprimat prin eroarea standard a mediei. Acesta, la rândul ei, este cu atât mai mare cu cât volumul eşantionului este mai mic. Desigur, cu cât limitele intervalului de estimare sunt mai apropiate de media eşantionului, cu atât aceasta din urmă estimează mai precis media populației şi prezintă mai multă încredere.

Testul *t* (Student) pentru un singur eşantion

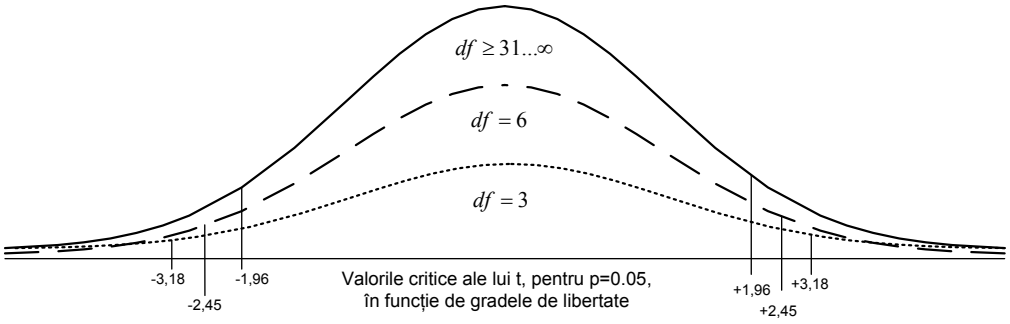
Aşa cum am precizat mai sus, **testul *z*** poate fi utilizat doar atunci când cunoaştem media populației de referință şi avem la dispoziție un eşantion „mare" (adică de cel puțin 30 de subiecți, în cazul unei variabile despre care avem motive să credem că se distribuie normal). Dar nu întotdeauna putem avea la dispoziție eşantioane „mari" (minim 30 de subiecți). Pentru situațiile care nu corespund acestei condiții, **testul *z*** nu poate fi aplicat. Şi aceasta, pentru că distribuția mediei de eşantionare urmează legea curbei normale standardizate doar pentru eşantioane de minim 30 de subiecți, conform teoremei limitei centrale.

La începutul secolului XX, William Gosset, angajat al unei companii producătoare de bere din SUA, trebuia să testeze calitatea unor eşantioane de bere pentru a trage concluzii asupra întregii şarje. Din considerente practice, el nu putea utiliza decât eşantioane (cantități) mici de bere. Pentru a rezolva problema,

Copyright © DEPARTAMENT ID 2008

a dezvoltat un model teoretic propriu, bazat pe un tip special de distribuție, denumită distribuție ***t***, cunoscută însă şi ca distribuția „Student", după pseudonimul cu care a semnat articolul în care şi-a expus modelul.

În esență, distribuția ***t*** este o distribuție teoretică care are toate caracteristicile unei distribuții normale (este perfect simetrică şi are formă de clopot). Specificul acestei distribuții constă în faptul că forma ei (mai exact, înălțimea) depinde de un parametru denumit „grade de libertate" (df sau *degrees of freedom*), care este egal cu N-1 (unde N este volumul eşantionului). Acest parametru poate fi orice număr mai mare decât 0, iar mărimea lui este aceea care defineşte forma exactă a curbei şi, implicit, proporția valorilor de sub curbă între diferite puncte ale acesteia. Imaginea de mai jos ilustrează modul de variație a înălțimii distribuției ***t***, în funcție de gradele de libertate.



Aşa cum se observă, curba devine din ce în ce mai aplatizată pe măsură ce **df** (volumul eşantionului) este mai mic. Acest fapt are drept consecință existența unui număr mai mare de valori spre extremele distribuției. Nu este însă greu de observat că, pe măsură ce **df** este mai mare, distribuția ***t*** se apropie de o distribuție normală standard astfel încât, pentru valori ale lui N de peste 31 (df=30), aria de sub curba distribuției ***t*** se apropie foarte mult de valorile de sub aria curbei normale standard (***z***), iar scorul critic pentru ***t*** este acelaşi ca şi cel pentru ***z*** pe curba normală (1.96).

Din cele spuse rezultă că, dacă avem un eşantion de volum mic (N<30), vom utiliza testul ***t*** în loc de testul ***z***, pe baza unei formule asemănătoare:

$$t = \frac{m - \mu}{s_m} \qquad \text{(formula 3.4)}$$

unde:

m este media eşantionului
μ este media populației
s_m este eroarea standard a mediei

Interpretarea valorii lui ***t*** se face în mod similar cu cea pentru valoarea lui *z*, cu deosebirea că se utilizează tabelul distributiei t (Anexa 2). În acest caz, valorile critice ale lui ***t*** vor fi diferite în funcție de numărul de grade de libertate. Citind tabelul, se observă că pragurile critice ale lui ***t*** (subînțelegând alfa=0.05, pentru test bilateral) se plasează la valori diferite în funcție de nivelul df. În acelaşi timp, dacă df este mare (peste 30), valorile tabelare ale lui ***t*** se apropie de cele ale lui ***z***. La infinit, ele sunt identice (±1.96, la fel ca şi în cazul valorilor lui ***z***).

Date fiind caracteristicile enunțate, în practică, **testul *t*** se poate utiliza şi pentru eşantioane mari (N≥30). În nici un caz însă, nu poate fi utilizat **testul *z*** pentru eşantioane mici (N<30). Utilizarea testului bazat pe un singur eşantion (fie ***z*** sau ***t***) depinde într-o măsură decisivă de asigurarea caracteristicii aleatoare a eşantionului.

Publicarea rezultatelor testului *z* sau t

Publicarea rezultatelor diferitelor proceduri statistice trebuie făcută astfel încât cititorii să îşi poată face o imagine corectă şi completă asupra rezultatelor. În acest scop la publicarea rezultatelor trebuie respectate anumite reguli, la care vom face trimitere în continuare, în legătură cu fiecare nou test statistic ce va fi introdus.

În principiu, publicarea rezultatelor unui test statistic se poate face în două moduri:

- sintetic (de regulă sub formă tabelară), atunci când numărul variabilelor testate este relativ mare;

Copyright © DEPARTAMENT ID 2008

- narativ, atunci când se referă, să zicem, la o singură variabilă.

În cazul testului pentru un singur eşantion se vor raporta: media eşantionului, media populației, valoarea lui ***z*** (sau ***t***), nivelul lui *p*, tipul de test (unilateral/bilateral).

Dacă avem în vedere rezultatele obținute pe exemplul de mai sus, se apelează la o raportare de tip narativ, care poate utiliza o formulare în maniera următoare: „*Eşantionul de elevi participanți la olimpiade a obținut un scor (QI=106; 95%CI:100.6-111.3) peste media populației generale (QI=100). Testul z, cu alfa 0.05, a demonstrat că diferența nu este semnificativă statistic, z=+2.13, p>0.05, unilateral*”.

În acest exemplu de prezentare nu formularea ca atare este esențială, ci informațiile asociate publicării testului ***z***. Formularea poate diferi de cea enunțată, dar elementele informaționale trebuie să fie complete. Expresia „*95%CI*” vine de la *95% Confidence Interval* şi exprimă intervalul de încredere pentru media populației.

Aşa cum am spus mai sus, utilizarea programelor statistice oferă pentru orice valoare a lui ***z*** (sau oricare alt test statistic) valoarea exactă a lui *p*. Ea poate fi utilizată ca atare, păstrând însă raportarea acesteia la pragul de semnificație. Orice valoare a lui *p* mai mare de 0.05 este considerată nesemnificativă[7], dacă nu a fost fixat un alt prag, mai sever.

Sarcină de lucru nr. 2. 2

La o evaluare a cunoştințelor de statistică indicatorii statistici descriptivi pentru studenții din întregul an de studiu (populația) sunt următorii:

μ=19.8
σ=3.91
N=192

Aceiaşi indicatori, pentru două cele două grupe de studiu care compun anul respectiv, sunt următoarele:

GR.	m	S	N
1	18.36	4.45	32
2	20.21	3.09	31

Presupunem că scorul obținut de dvs. la acest test de cunoştințe este 19.
Folosind indicatorii populației, ai grupei pe care ați ales-o şi scorul dvs. personal, calculați:

5. scorul ***z*** personal, raportat
    b. la fiecare dintre cele două grupe
    c. la întregul an
6. care este procentul valorilor mai bune decât scorul personal obținut, prin raportare la curba normală, pentru corul z calculat prin raportare la întregul an?
7. scorul ***z*** al fiecărei grupe, raportat la întregul an
8. care este procentul valorilor care se plasează pe curba normală peste scorul obținut de fiecare grupă?
9. Calculați limitele de încredere pentru media fiecărei grupe estimarea mediei întregului an, pentru ***z*** critic=1.85

Verificați răspunsurile corecte

[7] Programele de prelucrări statistice utilizează termenul „Sig.” (de la „*significance*” în loc de „*p*”. Ele sunt strict echivalente.

Copyright © DEPARTAMENT ID 2008

Erori statistice; Puterea testului statistic; Mărimea efectului

Să presupunem că avem un munte în care bănuim să se află aur (populația) şi ca dorim să dovedim prezența lui pe baza unei cantități de pământ extrase dintr-un loc ales la întâmplare (eşantion) din acest munte. Ipoteza de nul în acest caz afirmă că aurul nu este prezent în acest munte mai mult decât în orice alt loc. Mai departe, determinăm cantitatea de aur din eşantionul recoltat şi descoperim o anumită concentrație de metal prețios. În final, trebuie să hotărâm dacă această concentrație diferă de concentrația „naturală", pe care ne putem aştepta să o găsim oriunde. Dacă nivelul concentrației de aur din eşantion este mai mare decât cel al concentrației pe care ne aşteptăm să găsim în cel mult 5% (pragul alfa) din eşantioanele recoltate „din orice loc de pe pământ, ales la întâmplare", atunci suntem îndreptățiți să concluzionăm că aurul din eşantionul cercetării nu este „întâmplător" (respingem H0) şi, implicit, că „foarte probabil" muntele nostru conține aur într-o concentrație mai mare decât cea naturală (acceptăm H1).

Am spus mai sus „foarte probabil", fiindcă este evident faptul că nu putem fi absolut siguri de rezultatul nostru. În conformitate cu legea distribuției normale, dacă am recolta la întâmplare eşantioane de pământ, ne putem aştepta să avem situații în care concentrația de aur să fie oricât de mare, fără ca acest lucru sa însemne neapărat că „muntele" (populația cercetării) este un zăcământ aurifer (poate exista doar o zonă limitată, cu concentrație mare, iar restul muntelui să nu conțină aur). Aceasta înseamnă că asumarea deciziei cu privire la ipoteza de nul presupune implicit asumarea riscului unei anumite erori. Chiar dacă respectăm rigorile raționamentului şi deciziei statistice, nu avem garanția că decizia noastră reflectă „realitatea vieții". Cercetările statistice au un caracter probabilist şi, ca atare, conțin o anumită cantitate de eroare.

Erori statistice

În raport cu „realitatea vieții", decizia cu privire la ipoteza de nul poate fi corectă sau greşită dar, din păcate, cercetătorul care a efectuat studiul privind inteligența elevilor olimpici nu are cum să ştie cu certitudine dacă decizia pe care o ia este cu adevărat corectă sau este greşită. O imagine sintetică, frecvent utilizată pentru a ilustra relațiile posibile între decizia statistică şi „adevărul vieții", este prezentată în mod clasic prin următorul tablou:

		„Adevărul vieții" (necunoscut)	
		H_0 este adevărată (olimpicii NU SUNT mai inteligenți)	**H_0 este falsă** (olimpicii SUNT mai inteligenți)
Decizia statistică	**Acceptarea H_0** (olimpicii NU SUNT mai inteligenți)	1. decizie corectă p=1-*alfa*	*4.* ***eroare de tip II*** p=*beta*
	Respingerea H_0 (olimpicii SUNT mai inteligenți)	*2.* ***eroare de tip I*** P=*alfa*	3. decizie corectă p=1-*beta* ***(power)***

Aşa cum observăm, decizia statistică este corectă în două din celulele tabelului de mai sus: celula 1, acceptarea ipotezei de nul când ea este şi în realitate adevărată, şi celula 3, respingerea ipotezei de nul atunci când ea este şi în realitate falsă. În acest din urmă caz ne plasăm într-o situație statistică „ideală", în care decizia confirmă ipoteza cercetării, atunci când aceasta este adevărată şi în viața reală. Capacitatea unui test statistic de a susține o astfel de decizie, se numeşte „puterea testului statistic" (sau „puterea cercetării"), pe care o vom analiza pe larg puțin mai târziu. La rândul lor, erorile sunt ilustrate în celelalte două celule: celula 2, când respingem, ipoteza de nul, deşi ea este adevărată şi celula 4, când acceptăm ipoteza de nul, deşi ea este falsă. Pentru început, vom detalia situațiile de eroare.

În continuare, vom analiza în detaliu situațiile de eroare statistică.

Copyright © DEPARTAMENT ID 2008

Eroarea de tip I

Cercetătorul ştie că, chiar şi în cazul în care testul diferenței dintre media eşantionului şi media populației este mai mare decât valoarea critică corespunzătoare lui *alfa*, hazardul ar putea produce o diferență chiar mai mare decât cea constatată, fără nicio legătură cu prezența la olimpiadă. Rezultă de aici că, dacă pe baza rezultatului la testul statistic respingem ipoteza de nul şi acceptăm că participarea la olimpiade se asociază cu un nivel mai ridicat al inteligenței, o facem asumându-ne conştient riscul unei erori. Dacă diferența dintre cele două medii rezultă a fi semnificativă şi respingem ipoteza de nul, deşi conform „adevărului vieții" ea este adevărată, se comite o *eroare de tip I*. Probabilitatea acesteia este egală cu valoarea pragului *alfa*, al cărui nivel maxim acceptabil este fixat convențional la 0.05. Atunci când fixăm valoarea lui *alfa* (0.05 sau mai mică) drept criteriu de respingere a ipotezei de nul, definim, de fapt, cantitatea de eroare pe care suntem dispuşi să ne-o asumăm în a respinge ipoteza de nul, chiar dacă în realitate aceasta ar putea fi adevărată. Altfel spus, riscul de a decide că muntele conține un zăcământ aurifer, când de fapt acest lucru nu este adevărat. Din acest motiv, eroarea de tip I se concretizează într-un rezultat fals pozitiv.

Decizia statistică se bazează pe măsura în care eşantionul reprezintă în mod rezonabil caracteristicile populației. Chiar dacă selecția eşantionului s-a făcut în condiții ideale, există o anumită probabilitate (cu atât mai mare cu cât eşantionul este mai mic) ca valorile sale să se abată de la parametrii populației („adevărul vieții"). Ca urmare, putem să ne imaginăm o situație în care, chiar şi un eşantion selecționat aleatoriu să prezinte valori neobişnuit de îndepărtate de parametrii populației, fără nici o legătură cu condiția cercetării. Într-o astfel de situație, supunându-ne în mod corect regulilor convenționale ale deciziei statistice, respingem ipoteza de nul, făcând o *eroare de tip I* şi asumându-ne un rezultat fals pozitiv. Desigur, putem reduce probabilitatea *erorii de tip I* prin asumarea unei valori mai mici pentru *alfa* dar, aşa cum vom vedea mai departe, acest lucru nu este lipsit de consecințe.

Dacă privim în cvadrantul 1 din tabelul de mai sus, vom observa că probabilitatea de a decide corect, prin acceptarea ipotezei de nul atunci când ea este într-adevăr adevărată este egală cu 1-*alfa*. Acest lucru înseamnă că prin asumarea unei valori alfa=0.05, de exemplu, avem o probabilitate de 0.95 (1-0.5) de a accepta H0 când aceasta este în mod real adevărată. Din acest motiv valoarea din cadranul 1 se numeşte *nivel de încredere*. Ca să înțelegem şi mai bine, să ne imaginăm că am efectua exact acelaşi studiu de 100 de ori, utilizând eşantioane diferite, dar similare sub aspectul vârstei copiilor, volumului grupurilor şi procedurii etc. În cazul unei decizii statistice care respectă criteriile impuse, cu *alfa*=0.05 (implicit, 1-*alfa*=0.95), ne putem aştepta ca în 5% dintre aceste cercetări (100x0.05) să respingem în mod greşit ipoteza de nul (aceasta fiind, în realitate, adevărată). Acest lucru este echivalent cu a spune că avem un nivel de încredere de 95% (100x0.95) să acceptăm corect ipoteza de nul, dar şi că avem 95% şanse să acceptăm o ipoteză de nul care este în realitate adevărată. Cu alte cuvinte, valoarea lui *alfa* ne spune care este probabilitatea de a respinge în mod nejustificat o ipoteză de nul, adevărată în viața reală, eroare pe care însă cercetătorul este dispus să o tolereze.

Eroarea de tip II

Dar dacă, deşi muntele la care am făcut referire conține în mod real un zăcământ de aur, iar eşantionul nostru nu conține dovada acestui fapt şi ne sileşte să admitem ipoteza de nul? În acest caz comitem o eroare de tip II, care descrie un rezultat fals negativ.

Să presupunem că participarea la olimpiadă este asociată în mod real cu un nivel de inteligență mai ridicat dar, ca urmare a hazardului eşantionării, diferența dintre media eşantionului cercetării şi media populației nu atinge pragul semnificației statistice. Aceasta este situația în care, deşi elevii olimpici sunt mai inteligenți, cercetarea noastră are un rezultat nesemnificativ. Să nu uităm că cercetătorul nu cunoaşte care este „adevărul vieții" (dacă olimpicii sunt mai inteligenți) şi, drept urmare, chiar şi atunci când admite o ipoteză de nul îşi asumă un risc de eroare. Aceasta este o *eroare de tip II*, codificată cu *beta*. Admiterea existenței erorii de tip II nu este lipsită de controverse. Fisher, unul dintre teoreticienii marcanți ai statisticii moderne, considera că atunci când nu decidem respingerea ipotezei de nul, nu decidem acceptarea ei, ci doar consemnăm „eşecul de a o respinge", ceea ce nu este propriu-zis o decizie. Abia mai târziu, Neyman şi Egon Pearson (fiul lui Karl Pearson, autorul coeficientului de corelație care îi poartă numele) au dezvoltat teoria modernă a deciziei statistice, în prezent larg acceptată de comunitatea ştiințifică (B. Cohen, 2001).

Stabilirea nivelului probabilității erorii de tip II nu este uşor de înțeles, mai ales că ea este în legătură cu puterea testului, probabilitatea deciziei corecte, fixată în cadranul 3 al tabelului. Aceste două valori sunt complementare, puterea testului fiind egală cu 1-*beta*. În general, o valoare acceptabilă pentru eroarea de tip

Copyright © DEPARTAMENT ID 2008

II este *beta*=0.20, deoarece, aşa cum vom vedea mai târziu, valoarea recomandabilă pentru puterea testului este 0.80.

Atunci când inițiază studiul privind relația dintre inteligență şi participarea la olimpiadele şcolare, cercetătorul este interesat mai ales să evite admiterea ipotezei de nul atunci când aceasta ar fi, în realitate, falsă. Altfel spus, cercetătorul este interesat cu precădere în asumarea unei valori cât mai mici pentru *eroarea de tip II* (evitarea acceptării ipotezei de nul când ea este falsă), deoarece ar însemna că nu poate confirma ipoteza a cercetării. Micşorarea *erorii de tip II* ar însemna însă asumarea implicită a unei valori mai mari pentru riscul *erorii de tip I*. Se poate stabili o ierarhie între cele două tipuri de eroare? Este una mai „periculoasă decât alta? În mod obişnuit, „societatea" îşi impune punctul de vedere, declarând *eroarea de tip I* ca fiind mai „periculoasă", prin fixarea limitei maxime pentru *eroarea de tip I* (*alfa*=0.05). Dar de ce ar fi admiterea greşită a ipotezei de nul mai „rea" decât respingerea ei greşită? Aici trebuie să fim în consens cu Hack (2004) care afirmă că, deşi există o tendință de considerare a *erorii de tip I* ca fiind mai „rea" decât *eroarea de tip II*, în realitate ambele tipuri de erori pot fi la fel de „rele", prin consecințele practice care decurg din rezultatele cercetării.

Nu avem nici un motiv să credem că vreunul dintre cele două tipuri de eroare este mai „rău" sau mai „bun" decât celălalt. Dacă avem în vedere un criteriu moral, înainte de toate ar trebui să nu ne asumăm un rezultat pozitiv al cercetării, fără ca acest lucru să fie adevărat. Pe de altă parte, respingerea unui adevăr ştiințific numai pentru că cercetarea nu a fost în măsură să aducă dovada acestuia, este de asemenea de nedorit. Dacă am concluziona că muntele conține un zăcământ de aur, iar acest lucru s-ar dovedi fals, eroare de tip I, ar rezulta pierderi mari de organizare a unei exploatări ineficiente. La rândul ei, o eroare de tip II, care presupune admiterea ipotezei de nul şi negarea existenței unui zăcământ real, ar conduce la pierderi prin neexploatarea aurului existent. La fel, în plan psihologic, dacă obiectul testului statistic ar fi efectul unei noi metode de tratament psihoterapeutic, este la fel de rău să fie acceptată utilizarea ei, deşi nu este eficientă (*eroare de tip I*), ca şi respingerea utilizării, dacă ar fi eficientă (*eroare de tip II*), deoarece pacienții sunt lipsiți de un serviciu util.

Eroarea de tip III

Erorile de tip I şi II nu epuizează toate situațiile de eroare posibile într-o cercetare statistică. Howard Raiffa, într-o lucrare clasică de teoria deciziei, a introdus noțiunea de *eroare de tip III* (Raiffa, 1968). Ulterior, acest tip de eroare a fost luat în discuție şi de alți autori (Hack, 2004; Hsu, 1999), conturându-se două accepțiuni de bază ale termenului:

a. Respingerea corectă a ipotezei de nul, urmată de atribuirea incorectă a cauzei, definiție care corespunde cu definiția inițială propusă de Raiffa. În acest sens *eroarea de tip III* înseamnă o interpretare greşită a rezultatului. Cercetătorul concluzionează că „ceva semnificativ se întâmplă" şi, într-un fel, are dreptate, ceva se întâmplă, dar nu ceea ce crede el. Exemplul clasic este ilustrat de „efectul de noutate". Dacă introducem o noua metodă de antrenament bazată pe joc pentru stimularea învățării, copiii ar putea fi atraşi de noutatea situației în raport cu modalitatea clasică de învățare a regulilor de circulație. Ca urmare, un rezultat semnificativ diferit față de metoda utilizată pe un grup de control (care a învățat după metoda clasică) s-ar datora nu neapărat efectului noii metode, ci caracterului de noutate şi interes pe care îl prezintă aceasta. Este evident că cercetătorul este înclinat să considere efectul ca fiind generat de metoda investigată, dar acest lucru trebuie dovedit ca atare, nu este suficient să fie asumat. Efectul *placebo* poate fi inclus de asemenea în categoria erorilor de tip III, dar nu toate erorile de tip III sunt de tip *placebo*.

Nu există metode statistice pentru eliminarea *erorii de tip III*, în această accepție. Singura protecție vine dinspre calitatea modelului de cercetare. Pentru evaluarea efectului *placebo*, de exemplu, studiile medicale prevăd protocoale de tip „dublu orb", în care nici cei care administrează medicamentul şi nici pacienții nu ştiu dacă dau/iau medicamentul supus cercetării sau un *placebo*.

b. A doua definiție a *erorii de tip III* este similară cu prima, dar este diferită sub un aspect esențial. În acest caz rezultatul cercetării conduce la confirmarea unui „efect" sau „relații între variabile", dar sensul (direcția) efectului este greşit interpretat. Dacă revenim la exemplul anterior, ne putem imagina că rezultatele cercetării susțin concluzia că efectul noii metode de învățare este superior celei vechi deşi, în realitate, situația stă exact invers, concluzia fiind greşită. În această accepție, probabilitatea *erorii de tip III* este codificată cu litera γ (*gamma)*, iar unele programe statistice sunt capabile să o estimeze. Evident, eroarea de tip III se poate manifesta numai în cercetări de tip experimental, singurele care permit concluzii de natură cauzală.

Copyright © DEPARTAMENT ID 2008

Conceptul de *eroare de tip III* este fundamental diferit de celelalte două tipuri de erori. Existența lui vine să ne aducă aminte că cercetarea ştiințifică vizează în ultimă instanță un adevăr al realității, care nu este complet demonstrat de raționamentul decizional statistic, bazat pe atitudinea față de ipoteza cercetării şi admiterea sau respingerea ipotezei de nul. Principala lui utilitate este aceea că ne atrage atenția asupra vulnerabilității cercetărilor statistice, subliniind relativitatea acestora şi faptul că simpla declarare drept semnificativă a rezultatului unei cercetări nu probează în mod suficient adevărul ipotezei şi nici nu reflectă în mod sigur realitatea. Existența *erorii de tip III* este unul din argumentele împotriva asumării simpliste a rezultatelor statistice pe baza deciziei cu privire la ipoteza de nul. Mijlocul esențial de protecție împotriva *erorii de tip III* este stabilitatea rezultatelor de la o cercetare la alta, replicabilitatea lor, care înseamnă obținerea aceloraşi rezultate la repetarea studiului în aceleaşi condiții.

Puterea testului

Revenind la analogia cu muntele aurifer, să presupunem că rezultatul cercetării ne impune admiterea ipotezei de nul, implicit respingerea ipotezei că muntele conține aur. Într-un astfel de caz avem două posibilități de interpretare a acestui rezultat:

a. fie rezultatul cercetării este corect, ipoteza de nul este de fapt adevărată (ipoteza cercetării este realmente falsă), iar muntele nu conține aur (elevii olimpici nu sunt mai inteligenți decât populația elevilor în general);
b. fie ipoteza de nul este falsă, ceea ce ar însemna că zăcământul de aur există (olimpicii sunt mai inteligenți), dar explorarea noastră nu a avut suficientă „putere" („sensibilitate") pentru a surprinde existența aurului (relația dintre participarea la olimpiadă şi nivelul de inteligență). În acest caz, prin acceptarea ipotezei de nul (respingerea ipotezei cercetării) am comis o eroare de tip II.

„Puterea testului" este definită prin capacitatea sau „sensibilitatea" unui test statistic de a detecta un efect real (sau o legătură reală) între variabile. Înțelegem prin „efect real" faptul că modificări ale valorilor unei variabile se regăsesc în modificări ale valorilor celeilalte variabile (indiferent dacă relația este de tip cauzal sau de tip asociativ). Formulat în termeni statistici, puterea testului este probabilitatea de a respinge ipoteza de nul atunci când ea este cu adevărat falsă, şi se exprimă ca 1-*beta* (probabilitatea erorii de tip II). Această situație corespunde celei mai bune decizii pe care şi-o poate dori un cercetător: să dovedească că ipoteza a cercetării este realmente adevărată. Dacă în viața reală ipoteza de nul este falsă, dar datele cercetării ne obligă totuşi să o acceptăm, atunci putem spune că cercetarea noastră a avut o putere insuficientă pentru a determina respingerea ei şi, implicit, confirmarea ipotezei cercetării.

Aşa cum am văzut, *eroarea de tip II* şi puterea testului sunt complementare. Ca urmare, putem calcula *eroarea de tip II* ca *beta*=1-*puterea testului*. Cu alte cuvinte, cu cât puterea testului este mai mare, cu atât probabilitatea *erorii de tip II* (acceptarea nejustificată a ipotezei de nul) este mai mică. Dacă presupunem că puterea unui experiment psihologic este de 0.85, rezultă că probabilitatea erorii de tip II este 1-0.85, adică 0.15. Complementar, dacă puterea experimentului (cercetării) ar fi de 0.15, atunci probabilitatea *erorii de tip II* s-ar ridică la 1-0.15, adică 0.85.

Factori care contribuie la creşterea puterii testelor statistice

Puterea testului statistic sau, la fel de bine spus, a cercetării, poate fi calculată matematic. Introducerea procedurilor de calcul pentru puterea testului este dincolo de obiectivele pe care ni le propunem aici, mai ales că ele nu se regăsesc în pachetele obişnuite de analiză statistică. Vom reține însă, o serie de metode prin care poate fi asigurată creşterea puterii testelor statistice, aşa cum sunt ele sintetizate în literatura statistică (B. Cohen, 2004, Spata, 2003):

1. Aşa cum ştim, eroarea standard a mediei este cu atât mai mare cu cât eşantionul este mai mic. Ca urmare, una din modalitățile prin care putem creşte puterea este creşterea volumului eşantionului (N).
2. O cale de creştere a puterii este maximizarea variabilității primare, aceea care decurge ca urmare a „efectului" unei variabile asupra celeilalte. Aceasta deoarece „efectul" variabilei independente se manifestă mai puternic pe grupurile de subiecți aflate la extremitățile scalei de măsurare a variabilei dependente decât pe valorile întregii scale. Dacă împrăştierea datelor de cercetare este mică, atunci puterea testului de a surprinde un efect semnificativ se reduce.

Copyright © DEPARTAMENT ID 2008

3. Reducerea erorilor de măsurare are ca efect mărirea puterii cercetării. În acest scop trebuie avute în vedere: utilizarea unor proceduri de investigare adecvate; controlul şi eliminarea surselor de eroare; tratarea identică a tuturor subiecților cercetării; selectarea aleatoare a eşantioanelor sau, în cazul unei eşantionări nealeatoare, eliminarea surselor de selecție „părtinitoare" (*bias*).
4. Modelul de cercetare, prin el însuşi, este cel care poate creşte puterea unui studiu. De exemplu, modelele de cercetare *within-subjects* (intra-subiect), care măsoară aceiaşi subiecți în condiții diferite, au mai multă putere decât modelele *between-subjects* (inter-subiect), în care sunt comparate grupuri de subiecți diferiți în condiții diferite.
5. Testul bilateral reduce probabilitatea *erorii de tip I*, dar creşte probabilitatea *erorii de tip II* şi, implicit, reduce puterea. Ca urmare, ori de câte ori este justificabil, se va opta pentru test unilateral, chiar dacă, în practică, testul bilateral este cel uzual.
6. Testele parametrice prezintă o putere statistică mai mare decât cele neparametrice, motiv pentru care, utilizarea acestora din urmă se va face doar atunci când este absolut necesar (în conformitate cu condițiile de aplicare). Nu se va renunța cu uşurință la un test parametric, dacă datele cercetării sunt măsurate pe scală cantitativă.

Nu trebuie să înțelegem însă, că asigurarea unei puteri cât mai mari este principalul obiectiv pentru un cercetător. Prea multă putere este tot atât de nedorit ca şi prea puțină. Dacă avem în vedere intercondiționările din procesul deciziei statistice, atunci trebuie să observăm că prin creşterea puterii reducem probabilitatea *erorii de tip II*, dar creştem probabilitatea *erorii de tip I*. Cu alte cuvinte, dacă un studiu are o putere mare, de exemplu prin utilizarea unui eşantion foarte mare, atunci creşte probabilitatea de a respinge ipoteza de nul, chiar dacă aceasta este adevărată. Ne aflăm aici în situația care a generat critici vehemente cu privire la cercetările statistice, şi care a fost exprimată în maniera cea mai directă de Thompson (1998a) „*... testul statistic devine o căutare tautologică pentru suficienți participanți în măsură să atingă semnificația statistică*".

Calitatea deciziei unei cercetări reprezintă rezultatul unei „negocieri" între nivelul acceptat pentru erorile de tip I şi II. Cu cât prima este mai mică, cu atât a doua este mai mare, şi invers. Să presupunem că studiul privind inteligența olimpicilor este efectuat în mod identic de doi cercetători, dar unul dintre ei fixează nivelul lui alfa la 0.05, iar al doilea, la 0.01. Dacă în urma prelucrării datelor rezultatului obținut îi corespunde un p=0.03, primul cercetător va respinge ipoteza de nul, confirmând ipoteza cercetării, în timp ce al doilea va fi nevoit să admită ipoteza de nul şi să respingă ipoteza cercetării. Prin fixarea unui nivel mai redus pentru alfa, al doilea cercetător a redus probabilitatea *erorii de tip I*, dar a redus şi puterea testului, mărind în schimb riscul *erorii de tip II* (respingerea unei ipoteze de cercetare adevărate).

În concluzie, atunci când fixăm criteriile de decizie statistică trebuie să fim conştienți de următoarele aspecte:

- cu cât este mai mic pragul alfa, cu atât puterea testului este mai mică şi invers, cu cât alfa este mai mare, cu atât puterea testului este mai mare;
- cu cât alfa este mai mic, cu atât scade probabilitatea erorii de tip I (respingerea ipotezei de nul când aceasta este adevărată);
- cu cât alfa este mai mic, cu atât testul este mai „riguros", probabilitatea de a confirma ipoteza cercetării dacă este falsă, fiind mai mică;
- un prag alfa de 0.01 (comparat cu 0.05 sau 0.1) înseamnă că cercetătorul este precaut, dorind să îşi asume un risc de a greşi de 1 dintr-o sută de cazuri atunci când respinge ipoteza de nul, dacă aceasta este adevărată;
- un prag alfa de 0.01 înseamnă că există 99% şanse de a decide că nu există diferențe atunci când acestea într-adevăr nu există;
- mărind nivelul lui alfa (de la 0.01 la 0.05 sau 0.1), creştem riscul de a face o eroare de tip I şi reducem riscul de a face o eroare de tip II, ceea ce înseamnă şi o reducere a rigorii testului;
- în egală măsură, dacă mărim pragul alfa, de la 0.01, la 0.05 sau 0.1, mărim puterea, deoarece creştem probabilitatea de respingere a ipotezei de nul (acceptând ipoteza cercetării), atunci când aceasta din urmă este adevărată (eroare de tip I);

Din cele spuse s-ar putea deduce că, dacă ne propunem cea mai mare valoare pentru puterea testului, atunci singura opțiune pe care o avem este să fixăm pragul alfa la nivelul maxim permis de convenția ştiințifică (0.05). În realitate, problema nu este atât de simplă, deoarece obiectivul unei cercetări nu se poate limita doar la atingerea pragului de semnificație. Aşa cum am văzut, acesta poate fi atins prin mărirea volumului eşantionului, iar simpla constatare a unui rezultat semnificativ nu ne spune nimic despre

Copyright © DEPARTAMENT ID 2008

intensitatea relației dintre variabilele studiate, despre importanța practică şi despre utilitatea rezultatului obținut.

Cunoaşterea puterii unei cercetări este utilă în două situații:

a. În faza premergătoare a unei cercetări estimarea puterii este utilă pentru a evalua şansa de a obține un rezultat semnificativ statistic în contextul unei cercetări. Dacă puterea estimată a testului este prea mică, devine lipsit de interes să angajăm eforturi şi costuri pentru conducerea acelei cercetări. Cât de mică poate fi puterea unei cercetări pentru a accepta efectuarea ei? La aceasta întrebare cei mai mulți cercetători consideră că 0.5 este prea puțin pentru a investi timp şi bani în efectuarea ei. O putere de 0.7, care corespunde unei probabilități de 0.3 pentru *eroarea de tip II*, este considerată ca fiind minimă, iar o putere de 0.8 este considerat cel mai bun compromis între nivelul puterii şi consecințele negative de care am vorbit anterior (B. Cohen, 2001).

b. După efectuarea unei cercetări, pentru a şti care este probabilitatea ca rezultatul acesteia să indice un „efect" al variabilei independente asupra variabilei dependente atunci când acest efect există şi în realitate.

În practică calcularea puterii unei cercetări se face cu programe specializate. Unul dintre cele mai accesibile şi mai cunoscut dintre acestea este GPower, care poate fi descărcat gratuit de la adresa http://www.psycho.uni-duesseldorf.de/aap/projects/gpower/ (Buchner, Erdfelder & Faul, 1997).

Mărimea efectului

Să considerăm că rezultatul explorării muntelui presupus aurifer conduce la respingerea ipotezei de nul, iar geologii concluzionează că eşantionul conține aur într-o proporție „semnificativă". Înseamnă oare acest lucru că muntele conține „mult aur"? Desigur, nu. Înseamnă doar că acea cantitate de aur găsită în eşantion are o probabilitate prea mică să fie acolo din întâmplare, motiv pentru care s-a decis că prezența ei semnalează o concentrație „similară" la nivelul întregului munte (populații). Cât de „mare" este cantitatea de aur nu putem şti doar pe baza testului de semnificație statistică, deoarece acesta nu exprimă decât o decizie probabilistică şi nu o evaluare cantitativă.

Situația este identică în cazul cercetării cu privire la relația dintre participarea la olimpiadele şcolare şi nivelul de inteligență, unde am obținut pentru eşantionul de olimpici o medie QI=106. Aplicând criteriile deciziei statistice, am concluzionat că diferența de 6 unități față de media populației (QI=100) este semnificativă şi am respins ipoteza de nul. Dar ce putem spune despre această diferență, cât de „mare" este ea? În vorbirea curentă, prin „semnificativ" se înțelege şi „important" sau „mare". În cazul deciziei statistice însă, „semnificativ" are un înțeles limitat la expresia „*probabilitate prea mică pentru a rezulta din întâmplare*". De aceea, din ce în ce mai mulți autori (Daniel, 1998; Denis, 2003; Fan, 2001; Kotrlik & Williams, 2003; Thompson, 1998b) consideră că decizia statistică nu este suficientă pentru a proba integral valoarea unei ipoteze de cercetare. Respingerea ipotezei de nul pe baza criteriului *alfa* nu oferă suficientă informație cu privire la relația dintre variabilele cercetării. Este evident că rezultatul testului (QI=106) conține şi o componentă de „mărime". Dacă media eşantionului ar fi fost 108, sau 120, diferența ar fi fost mai mare decât 106. Şi totuşi, respingerea ipotezei de nul şi considerarea rezultatului drept „semnificativ" nu exprimă în nici un fel nivelul de „mărime" al diferenței. Mai mult, ne amintim că puterea testului creşte pe măsură ce creşte volumul eşantionului. Ca urmare, un rezultat „semnificativ" poate fi obținut fie şi numai prin creşterea numărului de subiecți, fără ca relația dintre cele două variabile să fie una „intensă".

Problema semnalată este mai acută decât pare la prima vedere. Criticii deciziei bazate pe testarea ipotezei de nul merg până acolo încât cer eliminarea acestui model de decizie cu privire la ipotezele cercetărilor ştiințifice. La rândul ei, *American Psychological Association* a organizat un grup de lucru având ca obiect elaborarea unor recomandări cu privire la raportarea rezultatelor statistice (Wilkinson&APA Task Force on Statistical Inference, 1999). Concluziile acestui grup de lucru stipulează că „*raportarea şi interpretarea mărimii efectului (...) este esențială pentru o cercetare bună*". În opinia autorilor, raportarea şi interpretarea mărimii efectului prezintă trei avantaje importante:

- facilitează studiile de metaanaliză (studii care sintetizează rezultatele mai multor cercetări pe aceeaşi temă);
- facilitează formularea unor ipoteze cu un grad mai mare de specificitate de către cercetătorii care vor studia aceeaşi temă;
- facilitează integrarea rezultatului unei cercetări în literatura dedicată acelui subiect,

Copyright © DEPARTAMENT ID 2008

Una dintre soluțiile acestei probleme este calcularea unui indice de „mărime a efectului" care oferă o informație suplimentară, extrem de utilă în interpretarea rezultatului testelor statistice. Această informație ne apropie mai mult de semnificația practică a rezultatului cercetării, ceea ce înseamnă mai mult decât semnificația statistică.

Calcularea mărimii efectului pentru testul *z* (*t*) pentru un singur eşantion

Indicele de mărime a efectului este, în esență, o valoare numerică ce exprimă „forța" sau „mărimea" relației dintre variabilele cercetate, indiferent dacă această este de tip cauzal sau nu. Principial, atunci când comparăm două medii, formula de calcul pentru mărimea efectului se bazează pe diferența dintre aceste medii, raportată la un indicator al variabilității.

În cazul testului z sau t pentru diferența dintre media unui eşantion şi media populației, indicele de mărime a efectului se calculează după formula lui Cohen (1988):

$$d = \frac{m - \mu}{\sigma} \text{ (formula 3.5)}$$

unde:

m=media eşantionului
μ=media populației
σ=abaterea standard a populației (atunci când nu o cunoaştem, putem utiliza abatarea standard a eşantionului)

Ca urmare, mărimea efectului pentru rezultatul cercetării cu privire la relația dintre participarea la olimpiadele şcolare şi nivelul inteligenței este:

$$d = \frac{106 - 100}{15} = 0.4$$

Dat fiind faptul că *d* este calculat prin raportarea diferenței la abaterea standard, el este considerat un indice standardizat al mărimii efectului. Acesta se exprimă printr-un număr zecimal cuprins între 0 (efect nul) şi 1 (efect maxim). Valori mai mari de 1 pot fi obținute uneori, dar numai în cazuri extreme. Valorile mici exprimă un nivel redus al intensității relației dintre variabile (chiar dacă este semnificativă), în timp ce valorile mari indică o relație „intensă" (puternică).

Dar cum putem să interpretăm valoarea lui *d*? O valoare ca cea obținută în cercetarea noastră este „mare", sau „mică"? În cazul explorării zăcământului aurifer, geologii pot estima suficient de exact cantitatea de aur pe care o pot extrage din zăcământ, pornind de la concentrația de aur din eşantionul explorat. În general, evaluările mărimii efectului în mediul ingineresc sunt de aşteptat să fie mult mai mari decât cele din cercetările socio-umane. Spre deosebire de ştiințele naturii, în psihologie răspunsul la această întrebare nu este uşor de găsit. Ca urmare cercetătorii sunt îndreptățiți să dezvolte propriile repere de apreciere a mărimii efectului ca fiind „mici", „medii" sau „mari". În psihologie, interpretarea valorii lui *d* se face după un model propus de Cohen (op.cit.), care a devenit un standard preluat de toți cercetătorii, şi care fixează doar trei praguri de mărime:

D (Cohen)	0.20	efect mic
	0.50	efect mediu
	0.80	efect mare

În conformitate cu recomandările lui Cohen, d=0.8 este considerat un efect mare. Nu atât de mare încât să rezulte ca evident prin observație directă, dar suficient de mare pentru a exista o bună şansă de a fi găsit ca statistic semnificativ prin utilizarea unui eşantion format dintr-un număr relativ mic de subiecți. Prin contrast, d=0.2 este considerat un efect mic. Pentru valori mai reduse decât atât, inițierea unei cercetări nu se justifică.

Revenind la studiul din exemplul nostru, rezultatul obținut corespunde unui nivel moderat al mărimii efectului (d=0.4). Sau, altfel spus, diferența dintre media inteligenței elevilor olimpici şi populația de elevi

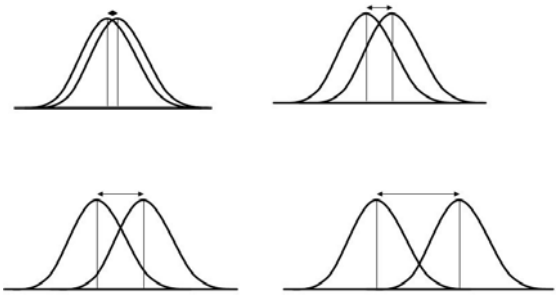
Copyright © DEPARTAMENT ID 2008

are un indice moderat de mărime. Acest lucru ar putea fi interpretat în sensul că prezența la olimpiadă este asociată în mod semnificativ cu inteligența, dar are şi alte componente importante care o determină.

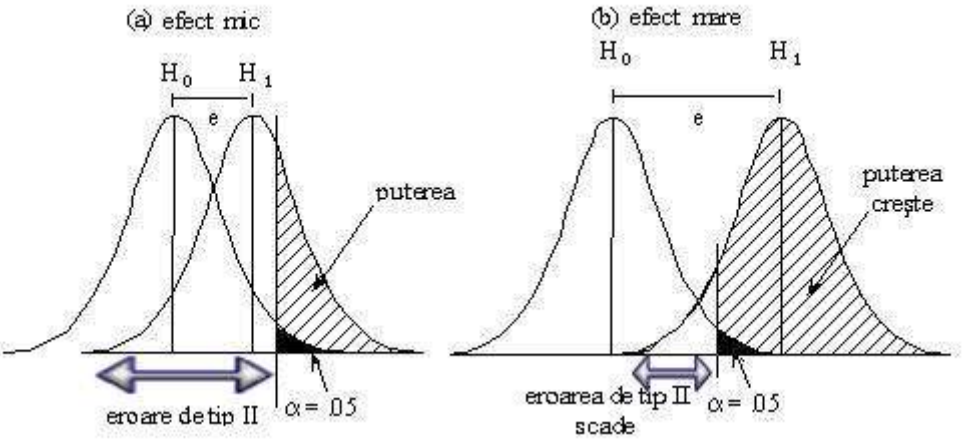
Calcularea mărimii efectului nu este oferită în toate situațiile de programele de prelucrare statistică. Din fericire, formulele de calcul nu sunt laborioase, putând fi aplicate cu uşurință pe rezultatele oferite de aceste programe. O prezentare sintetică şi practică a formulelor de calcul ale mărimii efectului pentru diverse teste statistice de semnificație ne oferă Thalheimer&Cook (2002).

Relația dintre mărimea efectului şi puterea testului

Mărimea efectului poate fi ilustrată prin gradul de suprapunere dintre distribuțiile supuse comparației (distribuția de nul şi distribuția cercetării). Cu cât suprafața comună a celor două distribuții este mai mică, mediile celor două distribuții devin tot mai îndepărtate una de alta, iar mărimea efectului creşte. Imaginea de mai jos ilustrează exact acest lucru:



În acelaşi timp, pe măsură ce creşte mărimea efectului, creşte şi puterea testului (concomitent cu reducerea riscului erorii de tip II):



Interpretare rezultatului unui test statistic

În contextul celor spuse până acum, pentru a putea interpreta mai complet rezultatele unei cercetări statistice, trebuie să ținem cont atât de nivelul de semnificație, cât şi de puterea testului şi de mărimea efectului. Un algoritm de evaluare a rezultatului la testul statistic este prezentat în tabloul următor:

Copyright © DEPARTAMENT ID 2008

Rezultat semnificativ statistic? (se respinge H0?)	Volumul eşantionului*	Concluzii
DA	**MIC**	• Rezultat important. • Chiar dacă puterea testului este mică, din cauza volumului redus al eşantionului, existența unui rezultat semnificativ arată o mărime a efectului importantă.
DA	**MARE**	• Rezultatul poate fi important sau nu: ➢ semnificația poate rezulta din puterea ridicată a testului, ca urmare a volumului mare al eşantionului SAU ➢ poate fi expresia unei diferențe importante dintre populațiile comparate
NU	**MIC**	• Rezultatul este neconcludent. Absența semnificației statistice se poate datora: ➢ faptului că ipoteza cercetării este falsă SAU ➢ puterii reduse a testului, ca urmare a eşantionului prea mică
NU	**MARE**	• Ipoteza cercetării este probabil falsă din cauză că, în ciuda puterii ridicate (eşantion mare), rezultatul nu a atins nivelul semnificației statistice. • Mărimea efectului este foarte mică

Rezumatul unității de învățare

- Scorul standard z exprimă distanța dintre o valoare a distribuției şi media acesteia, exprimată în abateri standard
- Media scorurilor z ale unei distribuții este întotdeauna egală cu 0.
- Abaterea standard a scorurilor z ale unei distribuții este întotdeauna egală cu 1.
- Alte tipuri de scoruri standard (QI, T, Hull, SAT, etc.) se calculează pe baza formulei scorului z, urmărind obținerea unei valori convenabile sub aspectul formei de expresie.
- Curba normală (Gauss) este o distribuție teoretică, caracteristic populațiilor de valori, care are o formă de clopot, este perfect simetrică şi asimptotică la axa Ox (poate lua, toretic, valori oricât de mari sau oricât de mici).
- Curba normală z, reprezintă o distribuție normală (Gauss) transformată în scoruri z. Aceasta poate exprima orice distribuție, indiferent de forma de exprimare a valorilor originale.
- Distribuțiile reale, transformate în distribuții z, nu îşi modifică forma originală.
- Distribuția de eşantionare este formată din totalitatea mediilor eşantioanelor de acelaşi volum posibile, extrase aleatori dintr-o populație.
- Media de eşantionare este egală cu media populației, dacă au fost extrase toate eşantioanele posibile.
- Eroarea standard a mediei este indicatorul de împrăştiere ale mediei de eşantionare.
- Eroarea standard a mediei este întotdeauna mai mică decât abaterea standard a populației.
- Teorema limitei centrale stipulează că distribuția de eşantionare tinde spre forma normală, atunci când eşantioanele extrase sunt suficient de mari (N este cel puțin 30).

* Facem precizarea că în acest context eşantion „mic" sau „mare" nu se referă la N=30 de subiecți la care am făcut referire în cazul teoremei limitei centrale. Se poate considera însă un eşantion „mic" ca fiind de ordinul zecilor, iar unul „mare" de ordinul sutelor.

Copyright © DEPARTAMENT ID 2008

- Scorul z pentru eşantion (grup), se calculează în acelaşi mod ca şi pentru valori individuale, cu deosebirea că în locul abaterii standard, se utilizează la numitor eroarea standard a mediei.
- Limitele de încredere ale mediei unui eşantion estimează, cu o anumită probabilitate, localizarea mediei la nivelul populației, în funcție de media eşantionului.
- Ipoteza cercetării descrie rezultatul aşteptat de cercetător la problema studiată.
- Ipoteza de nul reprezintă negația ipotezei cercetării şi face obiectul testării printr-o anumită procedură statistică.
- Decizia statistică este un raționament în baza căruia se admite sau se respinge ipoteza de nul.
- Pragul alfa este probabilitatea maximă ca rezultatul procedurii de tastare statistică să poată fi întâmplător, pe care o fixează cercetătorul drept criteriu de respingere sau de acceptare a ipotezei de nul.
- Pragul alfa=0.05 este nivelul maxim de probabilitate convențional acceptată de comunitatea ştiințifică pentru respingerea ipotezei de nul.
- Decizia unilaterală testează ipoteza statistică numai spre o latură a distribuției. Decizia bilaterală testează ipoteza în ambele direcții, cu menținerea pragului alfa stabilit.
- *Eroarea de tip I* este probabilitatea de a respinge o ipoteză de nul adevărată (se acceptă o ipoteză a cercetării care este falsă) – rezultat fals pozitiv.
- *Eroarea de tip II* este probabilitatea de a se admite o ipoteză de nul falsă (se respinge o ipoteză a cercetării adevărată) – rezultat fals negativ.
- O *eroare de tip III* apare atunci când rezultatul cercetării, deşi semnificativ, este greşit atribuit efectului variabilei independente, sau este în opoziție cu sensul real.
- Erorile de tip I şi II sunt în egală măsură negative dar, de regulă, acordăm mai multă atenție erorii de tip I, încercând să ținem alfa la o valoare cât mai mică.
- Puterea testului este o mărime probabilistă care indică şansa de a obține un rezultat semnificativ statistic.
- Puterea variază în funcție de nivelul pragului alfa (eroarea de tip I). Cu cât alfa este mai mic, cu atât puterea testului scade.
- Dacă reducem alfa de la 0.05 la 0.01, reducem probabilitatea de a face o eroare de tip I dar, în acelaşi timp, facem mai dificilă respingerea ipotezei de nul şi, în egală măsură, creştem probabilitatea de a face o eroare de tip II.
- Puterea testului este complementară erorii de tip II (suma lor este 1).
- Mărimea efectului este o valoare care indică intensitatea relației dintre variabila independentă şi variabila dependentă.
- Mărimea efectului este în legătură cu puterea testului şi cu volumul eşantionului. Cu cât puterea este mai mare şi eşantionul este mai mic, cu atât mărimea efectului este mai ridicată.
- Calcularea mărimii efectului, alături de semnificația statistică, este o exigență actuală în cercetarea ştiințifică psihologică.

Răspunsuri la sarcinile de lucru

Aten□ie, calculați utilizând numai primele două zecimale, fără rotunjiri. Eventuale mici varia□ii ale rezultatelor sunt posibile □i nu vor fi considerate erori.

Răspunsuri corecte la sarcina de lucru nr. 2. 1

Presupunem că evaluarea preferinței pentru risc la un grup de studenți aviatori care au suferit incidente critice în zbor a condus la o distribuție de valori având **m=60** şi **s=25**. Ştiind că indicele preferinței pentru risc la toți elevii piloți („populația") este **55**, şi are o distribuție normală, calculați răspunsul la următoarele întrebări:

1. z=(60-55)/(25/sqrt(30))=1,09

Copyright © DEPARTAMENT ID 2008

2. 36,21% (citit direct din tabelul distribu□iei z, la intersec□ia liniei z=1 cu coloana 0,09)
3. 50%-36,21%=13,79%
4. (50+36,21)=86,21% (procentul sub medie+procentul intre medie si z=1.09)
5. Calculam mai întâi z_{53}=(53-55)/(25/sqrt(30))=-0,65; Apoi căutăm în tabelul z probabilitatea la intersec□ia liniei 0,6 cu coloana 0,05 □i găsim p=0,24215. Această valoare reprezintă probabilitatea dintre medie (z=0) □i valoarea z=-0,65. Pentru a găsi probabilitatea unei valori mai mari de 53, adunăm probabilitatea valorilor peste medie (0.50) cu probabilitatea valorilor dintre -0,65 □i medie (0,24215)=**0,74215** (tabelul este simetric atit pentru valori pozitive cât □i pentru valori negative).
6. Calculăm z_{40}=(40-55)/(25/sqrt(30))=-1,09. Probabilitatea valorilor dintre mdie □i -1,09 este 0,36214., iar probabilitatea valorilor mai mici (adică mai îndepărtate de medie) este 50-0,36214=**0,13786**
7. Calculam mai întâi z_{45}=(45-55)/(25/sqrt(30))=-2,18 □i notăm probabilitatea asociată=0,48537; Calculăm apoi z_{48}==(48-55)/(25/sqrt(30))=-1,53 □i notăm probabilitatea asociată=-0,43699. Diferenta dintre aceste probabilită□i este răspunsul căutat: p=**0,04838**
8. Tabelul z ne da probabilită□ile dintre medie □i o anumită valoare z. ”Primii 10%” înseamnă ”cele mai mari 10% dintre valori”, ceea ce înseamnă că între acestea □i medie se află 40%. Citim celulele tabelului z pînă găsim cea mai apropiată cea mai apropiată valoare de 0,40 (0,39973). Apoi compunem scorul z limită din valorile liniei □i coloanei (z=1,28). În final, trebuie să transformăm scorul z în unită□ile de măsură ale scalei: X=55+1,28*25=**87** (am adunat fiindcă calculăm în dreapta mediei).
9. Tabelul z ne da probabilită□ile dintre medie □i o anumită valoare z. ”Ultimii 15%” înseamnă ”cele mai mici 15% dintre valori”, ceea ce înseamnă că între acestea □i medie se află 35%. Citim celulele tabelului z pînă găsim cea mai apropiată cea mai apropiată valoare de 0,35 (0,35083). Apoi compunem scorul z limită din valorile liniei □i coloanei (z=1,04). În final, trebuie să transformăm scorul z în unită□ile de măsură ale scalei: X=55-1,04*25=**29** (am scăzut fiindcă calculăm în stânga mediei).

Răspunsuri corecte la sarcina de lucru nr. 2. 2

1a. z(1)=0.14; z(2)=-031
1b. z(an)=-0.20
2. 57,93% (50% la care se adaugă 7,93% care corespunde ariei dintre medie şi scorul z=-0.20)
3. z(G1)=-2.08; z(G2)=0.58
4. Grupa 1: μ este cuprins între 7.92 şi 28,81; Grupa 2: μ este cuprins între 9,92 şi 30,49

Lucrarea de evaluare nr. 2.1

Lucrarea de evaluare va fi publicată pe portal (http://portal.credis.ro). Data limită de trimitere este preziua tutorialului. După acest termen lucrările nu mai sunt acceptate.

Copyright © DEPARTAMENT ID 2008

TESTE STATISTICE PARAMETRICE

Obiectivele unității de învățare şi informații introductive

Parcurgerea acestei unități va permite studenților:

- ✓ Să calculeze testele statistice z şi te pentru un singur eşantion
- ✓ Să calculeze şi să interpreteze semnificația diferenței dintre mediile a două eşantioane independente
- ✓ Să calculeze şi să interpreteze testul ANOVA pentru mai mult de două eşantioane independente
- ✓ Să calculeze şi să interpreteze testul t pentru eşantioane dependente
- ✓ Să calculeze şi să interpreteze testul de corelație liniară Pearson
- ✓ Să calculeze şi să interpreteze coeficientul de regresie liniară simplă

Testarea diferenței dintre mediile a două eşantioane independente

Testul ***z*** (**t**) pentru un singur eşantion este util într-un model de cercetare în care ne propunem compararea valorii măsurate pe un eşantion cu media populației din care acesta provine. Aşa cum am precizat deja, acest tip de cercetare este destul de rar întâlnit, ca urmare a dificultății de a avea acces la media populației.

Un model de cercetare mult mai frecvent însă, este acela care vizează punerea în evidență a diferențelor care există între două categorii de subiecți (diferența asumării riscului între bărbați şi femei, diferența dintre timpul de reacție al celor care au consumat o anumită cantitate de alcool față de al celor care nu au consumat alcool etc.). În situații de acest gen psihologul compară mediile unei variabile (preferința pentru risc, timpul de reacție etc.), măsurată pe două eşantioane compuse din subiecți care diferă sub aspectul unei alte variabile (sexul, consumul de alcool, etc.). Variabila supusă comparației este variabila dependentă, deoarece presupunem că suportă „efectul" variabilei sub care se disting cele două eşantioane şi care, din acest motiv, este variabilă independentă[8]. În studii de acest gen, eşantioanele supuse cercetării se numesc „independente", deoarece sunt constituite, fiecare, din subiecți diferiți.

Distribuția ipotezei de nul pentru diferența dintre medii independente

Să ne imaginăm că dorim să vedem dacă un lot de sportivi, trăgători la țintă, care practică *trainingul autogen*[9] (variabila independentă) obțin o performanță (variabila dependentă) mai bună decât un lot de sportivi care nu practică această tehnică de autocontrol psihic. În acest caz, variabila dependentă ia valori prin evaluarea performanței de tragere, iar variabila independentă ia valori convenționale, pe o scală nominală categorială, dihotomică (practicanți şi nepracticanți de şedințe de relaxare).

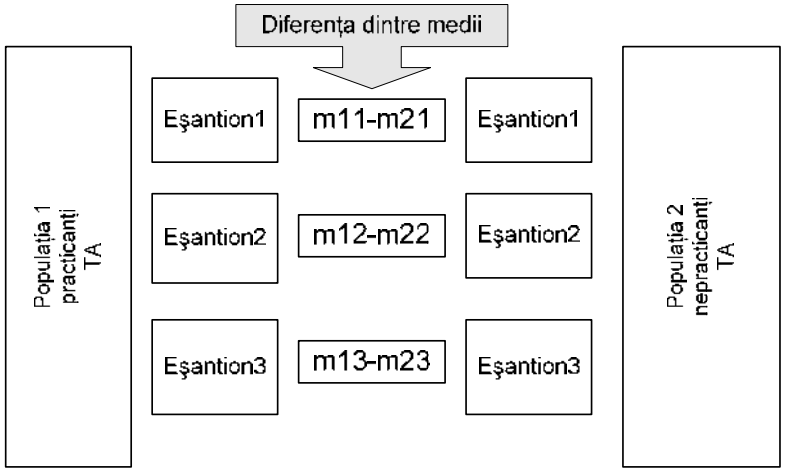
În acest exemplu avem două eşantioane de cercetare, unul format din sportivi practicanți ai trainingului autogen (TA) şi altul format din sportivi nepracticanți ai TA. Ipoteza cercetării susține că media performanței celor două grupuri este diferită. Sau, cu alte cuvinte, că cele două grupuri provin din populații diferite, respectiv, populația sportivilor practicanți de TA şi cea a nepracticanților de TA. Trebuie să

[8] Am pus cuvântul „efect" între ghilimele deoarece, chiar dacă este logic să considerăm că este vorba de o relație de tip cauză-efect, simpla măsurare a diferențelor pe două eşantioane de subiecți nu este suficientă pentru a concluziona o relație cauzală. Pentru aceasta, ar fi mai potrivit, spre exemplu, să măsurăm timpul de reacție la aceiaşi subiecți înainte şi după consumarea unei cantități de alcool.

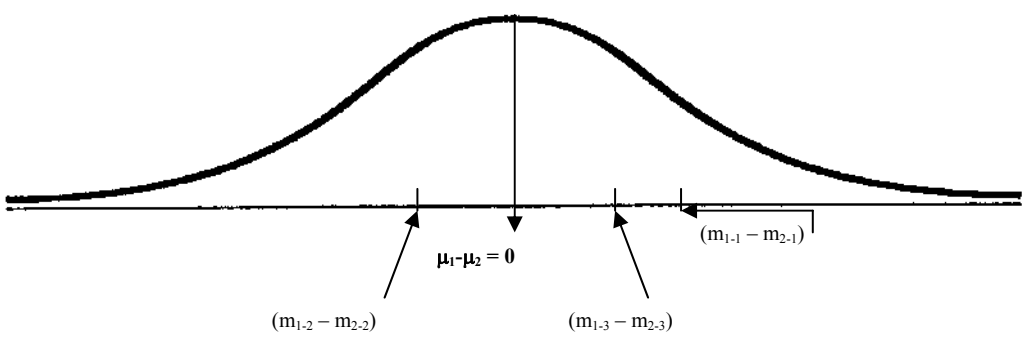
[9] O metodă de relaxare psihică

Copyright © DEPARTAMENT ID 2008

acceptăm faptul că perechea de eşantioane studiate nu este decât una din perechile posibile. Să privim figura de mai jos, care ne sugerează ce se întâmplă dacă, teoretic, am extrage (selecta) în mod repetat de eşantioane perechi din cele două populații:



Imaginea arată faptul că, pe măsură ce constituim perechi de eşantioane (*m11-m21*, etc.) cu valori ale performanței la țintă, diferența dintre medii devine o distribuție în sine, formată din valorile acestor diferențe. Dacă am reuşi constituirea tuturor perechilor posibile de eşantioane, această distribuție, la rândul ei, ar reprezenta o nouă populație, populația diferenței dintre mediile practicanților şi nepracticanților de training autogen. Şi, fapt important de reținut, curba diferențelor dintre medii urmează legea distribuției ***t***. Cu alte cuvinte, la un număr mare (tinzând spre infinit) de eşantioane perechi, trebuie să ne aşteptăm ca cele mai multe medii perechi sa fie apropiate ca valoare, diferența dintre mediile fiind, ca urmare, mică, tinzând spre 0 şi ocupând partea centrală a curbei. Diferențele din ce în ce mai mari fiind din ce în ce mai puțin probabile, vor ocupa marginile distribuției (vezi figura de mai jos). Aceasta este ceea ce se numeşte „distribuția ipotezei de nul" pentru diferența dintre mediile a două eşantioane independente.



În acest moment este bine să accentuăm din nou semnificația statistică a noțiunii de populație. După cum se observă, aceasta nu face referire neapărat la indivizi, ci la totalitatea valorilor posibile care descriu o anumită caracteristică (psihologică, biologică sau de altă natură). În cazul nostru, diferențele dintre mediile eşantioanelor perechi (fiecare provenind dintr-o „populație fizică" distinctă) devin o nouă „populație", de această dată statistică, compusă din totalitatea diferențelor posibile, a cărei distribuție se supune şi ea modelului curbei ***t***.

Procedura statistică pentru testarea semnificației diferenței dintre mediile a două eşantioane

Problema pe care trebuie să o rezolvăm este următoarea: este diferența dintre cele două eşantioane suficient de mare pentru a o putea considera că este în legătură cu variabila independentă, sau este doar una dintre diferențele probabile, generată de jocul hazardului la constituirea perechii de eşantioane? Vom observa că sarcina noastră se reduce, de fapt, la ceea ce am realizat anterior în cazul testului ***z*** sau **t** pentru un singur eşantion. Va trebui să vedem dacă diferența dintre două eşantioane reale se distanțează semnificativ de diferența la care ne putem aştepta în cazul extragerii absolut aleatoare a unor perechi de eşantioane, pentru care distribuția diferențelor este normală. Mai departe, dacă probabilitatea de a obține din întâmplare un

Copyright © DEPARTAMENT ID 2008

astfel de rezultat (diferență) este prea mică (maxim 5%) o putem neglija şi accepta ipoteza că între cele două variabile este o relație semnificativă.

Dacă avem valoarea diferenței dintre cele două eşantioane cercetate, ne mai sunt necesare doar media populației (de diferențe ale mediilor) şi abaterea standard a acesteia, pentru a calcula testul ***z*** (în cazul eşantioanelor mari) sau cel ***t*** (în cazul eşantioanelor mici). În final, nu ne rămâne decât să citim valoarea tabelară pentru a vedea care este probabilitatea de a se obține un rezultat mai bun (o diferență mai mare) pe o bază strict întâmplătoare.

Media populației de diferențe. Diferența dintre mediile celor două eşantioane ale cercetării face parte, aşa cum am spus, dintr-o populație compusă din toate diferențele posibile de eşantioane perechi. Media acestei populații este 0 (zero). Atunci când extragem un eşantion aleator dintr-o populație, valoarea sa tinde să se plaseze în zona centrala cea mai probabilă). Dar aceeaşi tendință o va avea şi media oricărui eşantion extras din populația pereche. Ca urmare, la calcularea diferenței dintre mediile a două eşantioane, cele mai probabile sunt diferențele mici, tinzând spre zero. Astfel, ele vor ocupa partea centrală a distribuției, conturând o medie tot mai aproape de zero cu cât numărul eşantioanelor extrase va fi mai mare.

Eroarea standard a diferenței (împrăştierea), pe care o vom nota cu σ_{m1-m2}, se calculează pornind de la formula de calcul a erorii standard:

$$\sigma_m = \frac{\sigma}{\sqrt{N}} \qquad \text{(formula 3.6)}$$

Din rațiuni practice, pentru a obține o formulă care să sugereze diferența dintre medii (m_1-m_2), formula de mai sus este supusă unor transformări succesive. Prin ridicarea la pătrat a ambilor termeni, şi după extragerea radicalului din noua expresie, se obține:

$$\sigma_m = \sqrt{\frac{\sigma^2}{N}} \qquad \text{(formula 3.7)}$$

Dacă am utiliza-o pentru calcule, această ultimă formulă ar produce acelaşi rezultat ca şi formula de origine.

Formula erorii standard a distribuției diferenței dintre medii ne arată cât de mare este împrăştierea diferenței „tipice" între două medii independente atunci când eşantioanele sunt extrase la întâmplare

$$\sigma_{m1-m2} = \sqrt{\frac{\sigma_1^2}{N_1} + \frac{\sigma_2^2}{N_2}} \qquad \text{(formula 3.8)}$$

Formula 3.8 ne indică faptul că eroarea standard a diferenței dintre medii este dată de suma erorii standard a celor două eşantioane. Unul dintre eşantioane are N_1 subiecți şi o dispersie σ_1^2 iar celălalt eşantion, N_2 subiecți şi dispersia σ_2^2. Faptul că obținem eroarea standard a <u>diferenței</u> dintre medii ca <u>sumă</u> a erorilor standard a celor două eşantioane este fundamentat pe o lege statistica a cărei demonstrație nu se justifică aici.

Pentru a calcula scorul ***z*** al diferenței, vom utiliza o formulă asemănătoare cu formula notei z pe care o cunoaştem deja:

$$z = \frac{m - \mu}{\sigma_m}$$

Aceasta va fi:

$$z = \frac{(m_1 - m_2) - (\mu_1 - \mu_2)}{\sigma_{m1-m2}} \qquad \text{(formula 3.9)}$$

<u>Numărătorul</u> exprimă diferența dintre diferența obținută de noi (m_1-m_2) şi diferența dintre mediile populațiilor (μ_1-μ_2). Dacă ne amintim că distribuția ipotezei de nul (μ_1-μ_2) are media 0, atunci deducem că expresia (μ_1-μ_2) poate lipsi. De altfel, dacă am cunoaşte mediile celor două populații nici nu ar mai fi necesară calcularea semnificației diferenței dintre eşantioanele care le reprezintă.

Copyright © DEPARTAMENT ID 2008

Numitorul descrie eroarea standard a diferenței, calculată cu formula 3.7, adică împrăştierea diferenței „tipice" pentru extrageri aleatoare.

În conformitate cu cele spuse până acum, formula finală pentru scorul ***z*** al diferenței dintre două eşantioane devine :

$$z = \frac{m_1 - m_2}{\sqrt{\frac{{\sigma_1}^2}{N_1} + \frac{{\sigma_2}^2}{N_2}}} \qquad \text{(formula 3.10)}$$

Se observă că am eliminat (μ_1-μ_2) de la numărător, care este întotdeauna 0 şi am înlocuit σ_{m1-m2} cu expresia echivalentă din formula 3.8. Această formulă ne dă ceea ce se numeşte valoarea *testului z pentru eşantioane mari-independente*.

Valoarea astfel obținută urmează a fi verificată cu ajutorul tabelei ***z*** pentru curba normală, iar decizia statistică se ia în acelaşi mod ca şi în cazul testului ***z*** pentru un singur eşantion.

În formula 3.9 eroarea standard a diferențelor este calculată pe baza erorii standard a distribuției de eşantionare pentru populațiile din care sunt extrase cele două eşantioane („practicanți" şi „nepracticanți" de training autogen). În realitate nu cunoaştem cele două dispersii. Din fericire, dacă volumul însumat (N_1+N_2) al eşantioanelor care dau diferența noastră (m_1-m_2) este suficient de mare (≥30 dar, de preferat, cât mai aproape de 100) atunci ne amintim că putem folosi abaterea standard a fiecărui eşantion (s_1 respectiv s_2), care aproximează suficient de bine abaterile standard ale celor două populații.

Atunci când eşantioanele nu sunt suficient de mari, trebuie să ne aşteptăm la erori considerabile în estimarea împrăştierii populației pe baza împrăştierii eşantionului. Într-o astfel de situație vom apela, desigur, la un test ***t***, având două opțiuni de calcularea acestuia:

a. Testul *t* pentru dispersii diferite

Acesta se bazează pe considerarea separată a dispersiilor celor două populații (estimate prin dispersiile eşantioanelor). Formula este foarte asemănătoare cu formula anterioară pentru testul ***z***. Vom reține această formulă ca **testul *t* pentru dispersii diferite:**

$$t = \frac{m_1 - m_2}{\sqrt{\frac{{s_1}^2}{N_1} + \frac{{s_2}^2}{N_2}}} \qquad \text{(formula 3.11)}$$

Se observă înlocuirea lui σ (pentru populație) cu s (pentru eşantion). Utilizarea acestei formule este destul de controversată deoarece rezultatul nu urmează cu exactitate distribuția ***t***, aşa cum am introdus-o anterior. Pentru eliminarea acestui neajuns, se utilizează o altă variantă de calcul, care ia în considerare dispersia cumulată a celor două eşantioane.

b. Testul *t* pentru dispersia cumulată

Dispersiile celor două eşantioane pot fi considerate împreună pentru a forma o singură estimare a dispersiei populației (σ^2). Obținem astfel ceea ce se numeşte „dispersia cumulată", pe care o vom nota cu s^2_c şi o vom calcula cu formula următoare:

$${s^2}_c = \frac{(N_1 - 1) * {s_1}^2 + (N_2 - 1) * {s_2}^2}{N_1 + N_2 - 2} \qquad \text{(formula 3.12)}$$

La numărător, formula conține suma dispersiilor multiplicate, fiecare, cu volumul eşantionului respectiv (de fapt, gradele de libertate, N-1). În acest fel vom avea o contribuție proporțională cu numărul de valori ale împrăştierii fiecărui eşantion la rezultatul final.

La numitor, avem gradele de libertate (*df*) pentru cele două eşantioane luate împreună (N_1+N_2-2).

Înlocuind-o în formula 3.11, obținem formula de calcul a testului ***t*** pentru dispersii cumulate:

Copyright © DEPARTAMENT ID 2008

$$t = \frac{m_1 - m_2}{\sqrt{\frac{(N_1 - 1) * {s_1}^2 + (N_2 - 1) * {s_2}^2}{N_1 + N_2 - 2} * \left(\frac{1}{N_1} + \frac{1}{N_2}\right)}}$$

(formula 3.13):

Expresia 3.13 este formula uzuală pentru calcularea diferenței dintre medii pentru două eşantioane independente. Chiar dacă a fost introdusă ca utilizabilă pentru „eşantioane mici", caracteristicile distribuției ***t*** ne permit utilizarea ei şi pentru eşantioane mari, deoarece distribuția ***t*** tinde spre cea normală la valori din ce în ce mai mari ale gradelor de libertate.

EXEMPLU DE CALCUL:

Să presupunem că vrem să vedem dacă practicarea trainingului autogen (variabila independentă) determină o creştere a performanței în tragerea la țintă, manifestată printr-un număr mai mare de lovituri în centru țintei (variabilă dependentă). Pentru aceasta selectăm un eşantion de 6 sportivi care practică trainingul autogen şi un eşantion de 6 sportivi care nu îl practică. Pentru fiecare eşantion măsurăm performanța de tragere.

Formularea ipotezei cercetării, a ipotezei de nul, şi a criteriilor deciziei statistice

Pentru exemplul de mai sus:

Problema cercetării: Are practicarea trainingului autogen un efect asupra performanței la tirul cu arcul?

Ipoteza cercetării (H_1): „*Practicarea trainingului autogen determină un număr mai mare de puncte la şedințele de tragere*".

Ipoteza de nul (statistică) (H_0): "*Numărul punctelor la şedințele de tragere nu este mai mare la cei care practică trainingul autogen*". Această variantă este potrivită cu o testare unilaterală a ipotezei (nu avem în vedere decât eventualitatea ca trainingul autogen să crească performanța sportivă).

Dacă, însă, am dori să testăm în ambele direcții, bilateral, atunci am avea următoarele versiuni ale ipotezelor:

Ipoteza cercetării: „*Performanța sportivă este diferită la subiecții care practică trainig autogen față de cei care nu practică*"

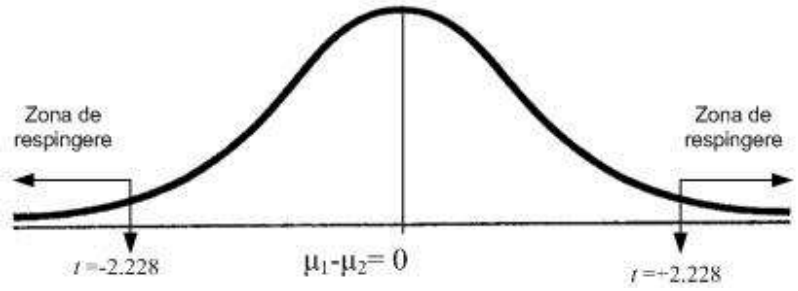
Ipoteza de nul (statistică): „*Performanța nu diferă semnificativ în funcție de practicarea trainingului autogen*".

*Fixarea lui **t** critic.* Optăm pentru efectuarea unui test bilateral, pentru că nu putem şti dinainte dacă TA nu are un efect negativ asupra performanței sportive a trăgătorilor la țintă. Alegem nivelul α=0,05. Stabilim gradele de libertate:

$df=N_1+N_2-2=10$

Utilizând tabelul distribuției ***t*** pentru 10 grade de libertate (adică 12-2) şi α=0,05, bilateral, găsim ***t*** critic=±2.228, la intersecția coloanei 0.025 şi cu linia pentru 10 grade de libertate.

Valoarea ***t*** calculată va trebui să fie cel puțin egală sau mai mare decât ***t*** critic, pentru a putea respinge ipoteza de nul şi a accepta ipoteza cercetării (vezi imaginea de mai jos).



Variabila independentă (calitatea de practicant-nepracticant Training Autogen) ia două valori, să zicem: „1" pentru practicanții trainingului autogen şi „2" pentru nepracticanți. Valorile „1" şi „2" sunt convenționale şi ne indică faptul că variabila independentă a cercetării noastre este măsurată pe o scală nominală, categorială (dihotomică). Variabila dependentă (performanța de tragere la țintă) ia valori cantitative, exprimată în număr de lovituri în centrul țintei, fiind de tip cantitativ (raport).

Copyright © DEPARTAMENT ID 2008

Datele cercetării

	practicanți TA („1")		ne-practicanți TA („2")	
	X1	**$(X1-m1)^2$**	**X2**	**$(X2-m2)^2$**
	15	2.78	10	2.78
	9	18.74	8	0.10
	12	1.76	11	7.12
	13	0.10	5	11.08
	16	7.12	7	1.76
	15	2.78	9	0.44
Σ	80	33.28	50	23.28
N	6		6	
m	13.33		8.33	
$s^2 = \frac{\sum (X_i - m)^2}{N-1}$	$\frac{33.28}{5} = 6.67$		$\frac{23.28}{5} = 4.66$	
$s = \sqrt{s^2}$	2.58		2.16	

Calculăm testul ***t*** pentru dispersii cumulate:
Mai întâi, eroarea standard a diferenței (numitorul formulei):

$$s_{Dif} = \sqrt{\frac{(N_1-1)(s^2{}_1)+(N_2-1)(s^2{}_2)}{N_1+N_2-2}\left(\frac{1}{N_1}+\frac{1}{N_2}\right)} = \sqrt{\frac{(6-1)*(2.58)^2+(6-1)*(2.16)^2}{10}\left(\frac{1}{6}+\frac{1}{6}\right)} = 1.34$$

Iar apoi:

$$t = \frac{m_1 - m_2}{s_{Dif}} = \frac{13.33 - 8.33}{1.34} = 3.73$$

Comparăm t calculat cu t critic din tabelul distribuției t: 3.73 > 2.228
Decizia statistică: Se respinge ipoteza de nul
Concluzia cercetării: Se admite ipoteza cercetării. „*Practicarea trainingului autogen este în legătură cu performanța de tragere*"

Mărimea efectului

Atunci când calculăm testul ***t***, nu valoarea obținută este relevantă ci probabilitatea care este asociată acestei valori (*p*). De exemplu, dacă avem în vedere formula de calcul pentru t, atunci înțelegem că o valoare t=3.73 nu înseamnă altceva decât faptul că diferența dintre mediile comparate este 3.73 ori mai mare decât eroarea standard estimată a acelei diferențe. Chiar dacă probabilitatea asociată acestei valori ***t*** este foarte mică, sub pragul alfa, magnitudinea diferenței dintre medii poate fi mică. Ca urmare, aprecierea „importanței" diferenței dintre mediile grupurilor cercetate are nevoie de informații suplimentare. Acestea sunt oferite de indicele de mărime a efectului.

Pentru a afla „mărimea efectului" pentru testul t pentru eşantioane independente, se utilizează indicele d al lui Cohen. Din păcate, pachetele de programe statistice uzuale (inclusiv SPSS) nu oferă acest valoarea lui d. El poate fi însă obținut relativ uşor cu formula:

$$d = \frac{m_1 - m_2}{\sqrt{\frac{(N_1-1)*s^2{}_1+(N_2-1)*s^2{}_2}{(N_1-1)+(N_2-1)}}} \quad \text{(formula 3.14)}$$

unde numitorul exprimă abatarea standard cumulată a celor două grupuri comparate.

Pentru exemplul nostru, calculăm mărimea efectului înlocuind datele în formula 3.14, după cum urmează:

Copyright © DEPARTAMENT ID 2008

$$d = \frac{13.33 - 8.33}{\sqrt{\frac{(6-1)*6.67 + (6-1)*4.66}{(6-1)+(6-1)}}} = \frac{5}{2.38} = 2.1$$

Interpretarea mărimii lui d se face utilizând aceleaşi praguri propuse de Cohen: 0.20 – efect mic; 0.50 – efect mediu; 0.80 – efect mare. Valoarea obținută de noi indică un nivel ridicat al mărimii efectului, semn al faptului că practicarea şedinţelor de relaxare are un „efect" important asupra performanței sportivilor din eşantionul cercetării.

Limitele de încredere ale diferenței dintre medii

Aşa cum ştim, mediile grupurilor comparate reprezintă doar o estimare a mediei populațiilor din care provin, oscilând jurul mediei „adevărate". În mod similar, diferența dintre mediile celor două eşantioane estimează media populației de diferențe. Cât de precisă este această estimare putem afla prin calcularea intervalului de încredere pentru diferența mediilor. Principial, limitele de încredere în acest caz se calculează la fel ca şi limitele de încredere pentru media populației, după următoarea formulă:

$$\mu_{dif} = m_{dif} \pm t_{critic} * s_{dif} \quad \text{(formula 3.15)}$$

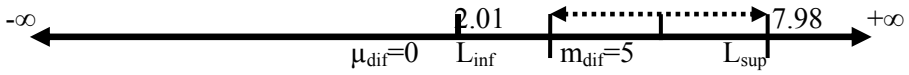
unde:

μ_{dif}=media populației de diferențe (μ_1-μ_2)
m_{dif}=diferența dintre mediile eşantioanelor cercetării (m_1-m_2)
t_{critic}=valoarea lui t pentru nivelul de încredere ales (de regulă 95%)
s_{dif}=eroarea standard a diferenței (calculată cu expresia de la numitorul formulei 3.13)

Înlocuind datele în formulă, obținem următoarele limite de încredere pentru media populației de diferențe:

Limita inferioară μ_{dif}=5-2.228*1.34=2.01
Limita superioară μ_{dif}=5+2.228*1.34=7.98

Imaginea de mai jos ilustrează limitele între care se află, pe distribuția populației de diferențe, având media 0, cu un nivel de încredere de 95%, poziția mediei reale a diferenței dintre grupurile comparate:



Relevanța intervalului de încredere poate fi discutată din mai multe puncte de vedere:

(a) Faptul că media populației de nul (μ_{dif}=0) se află în afara limitelor de încrerede subliniază odată în plus caracterul semnificativ al diferenței dintre mediile grupurilor comparate. Cu cât una dintre limite ar fi mai aproape de valoarea 0, cu atât faptul de a fi obținut un rezultat semnificativ ar fi mai puțin relevant. Dacă media distribuției de nul ar fi cuprinsă între limitele de încredere ipoteza de nul ar trebui acceptată, indiferent de rezultatul testului statistic.

(b) Mărimea intervalului de încredere arată precizia estimării rezultatului cercetării. Aceasta este legată în mod direct de eroarea standard a diferenței (eroarea de estimare) care, la rândul ei, depinde de numărul subiecților din cele două eşantioane, dar şi de omogenitatea valorilor măsurate.

(c) În măsura în care variabila testată are o utilitate practică, limitele de încredere scot în evidență dacă rezultatul are o semnificație în raport cu criterii de ordin practic. De exemplu, în cazul nostru, antrenorul sportivilor respectivi poate aprecia în ce măsură un progres al performanței care poate fi între 2 şi 7 puncte ar aduce o clasare mai bună la concursurile de profil sau, dimpotrivă, este „nerentabil".

(d) Limitele de încredere nu prezintă o utilitate practică atunci când valorile variabilei nu au o semnificație prin ele însele. Să ne imaginăm, spre exemplu, un experiment în care un grup priveşte un film trist, iar un alt grup priveşte un film vesel, după care starea de spirit a celor două grupuri este evaluată prin numărarea cuvintelor triste sau vesele pe care subiecții şi le pot aminti dintr-o listă citită imediat după vizionare. În această situație este greu de atribuit o utilitate practică limitelor de încredere ale „numărului de cuvinte evocate". Nu acelaşi lucru se întâmplă dacă, de exemplu, în cazul unui experiment în care utilizarea unui anumit tip de exerciții la locul de muncă se traduce în creşterea productivității muncii, măsurată prin numărul de produse finite. Este evident că numărul de produse finite este un indicator cu relevanță practică,

Copyright © DEPARTAMENT ID 2008

uşor de interpretat. Cu toate acestea, chiar şi atunci când nu prezintă o relevanţă practică directă, calcularea limitelor de încredere oferă o imagine a gradului de precizie a estimării testului statistic, fapt care face necesară cunoaşterea lor şi raportarea lor.

Interpretarea rezultatului la testul *t* pentru eşantioane independente

Atunci când valoarea calculată a testului este egală sau mai mare decât t critic (ceea ce este echivalent cu „*p* este mai mic sau egal cu alfa"), rezultatul justifică aprecierea ca semnificativă a diferenței dintre mediile celor două eşantioane (adică suficient de mare pentru a respinge ipoteza că ar putea fi întâmplătoare). Modelul de cercetare nu permite formularea acestei concluzii în termenii unei relații cauzale între practicarea trainingului autogen şi performanţa sportivă, oricât de tentată ar fi această concluzie. Cel puțin nu în contextul acestui model de de cercetare. Dacă acelaşi grup de subiecți ar fi fost supus evaluării performanței de extragere în zile cu training autogen şi în zile fără training autogen, concluzia ar fi putut fi de ordin cauzal.

În plus, existența unei diferențe semnificative nu este similară cu existența unei diferențe cu valoare practică. Este posibil ca diferența dintre cele două loturi de sportivi, deşi semnificativă statistic, să nu justifice costurile angajate în desfăşurarea programului de relaxare psihică. Într-o asemenea situație, studiul nu este lipsit de valoare dar concluziile sunt utile doar în plan teoretic.

Publicarea rezultatului

La publicarea testului ***t*** pentru diferența dintre mediile a două eşantioane independente vor fi menționate: mediile şi abaterile standard ale fiecărui eşantion, volumul eşantioanelor sau gradele de libertate, valoarea testului, nivelul lui *p*, mărimea efectului şi limitele de intervalului de încredere pentru diferența dintre medii.

În formă narativă, rezultatul pentru exemplul de mai sus poate fi formulat astfel: „*Sportivii care practică trainingul autogen au fost comparați cu cei care nu practică. Primii au realizat o performanță mai bună (m=13.33, σ=2.58) față de ceilalți (m=8.33, σ=2.16), t(10)=3.65, p<0.05. Mărimea efectului este mare (d=2.1), iar limitele de încredere (95%) pentru diferența mediilor sunt cuprinse între 2.01 şi 7.98*".

Condițiile în care putem calcula testul *t* pentru eşantioane independente

- Eşantioane aleatoare (ideal), sau neafectate de erori de eşantionare (*bias*);
- Eşantioane independente (distincte din punctul de vedere al variabilei independente, care determină constituirea grupurilor);
- Variabila supusă măsurării să se distribuie normal în ambele populații. Aceasta ne garantează că şi distribuția diferențelor dintre medii se distribuie normal. Totuşi, teorema limitei centrale ne permite asumarea normalității distribuției mediei de eşantionare chiar şi în cazul variabilelor care nu se distribuie normal la nivelul populației, pentru eşantioane mari. Dacă însă, analiza distribuțiilor indică forme aberante, iar volumul grupurilor comparate este foarte mic, se va alege soluția unui test neparametric. Vom menționa, totuşi, că testele ***t*** sunt robuste la încălcarea condiției de normalitate.
- Dispersia celor două eşantioane să fie omogenă. Testul **t** poate fi aplicat strict în cazurile în care dispersiile celor două populații („practicanți", „nepracticanți") au aceeaşi dispersie (omogenitatea dispersiei). Din fericire, există trei situații în care această condiție nu trebuie să ne preocupe:
    - când eşantioanele sunt suficient de mari (cel puțin 100 fiecare)
    - când cele două eşantioane au acelaşi volum (N1=N2)
    - când dispersiile celor două eşantioane nu diferă semnificativ (dar, chiar şi pentru acest caz, există formule care țin cont de diferența dispersiilor).

Când se utilizează testul *t* pentru eşantioane independente?

Generic, acest test statistic se utilizează în situațiile în care vrem sa aflăm dacă o variabilă dependentă, măsurată pe o scală de interval/raport, diferă semnificativ între două grupuri (eşantioane) diferențiate pe o variabilă independentă măsurată pe scala de tip nominal (dihotomic), sau bi-categorială, indiferent de natura ei. Deoarece este unul dintre modelele frecvent întâlnite în practica cercetării

Copyright © DEPARTAMENT ID 2008

psihologice, utilizarea testului ***t*** pentru eşantioane independente este şi ea des întâlnită în literatura de specialitate.

Sarcina de lucru nr. 2.4

Într-un studiu asupra efectelor unui nou tratament al fobiei, datele pentru grupul experimental obținute printr-o scală de evaluare a tendințelor fobice sunt: m_1=27.2, s_1=4 şi N_1=15

Datele pentru grupul de control sunt: m_2=34.4, s_2=14 şi N_2=15

Utilizând aceste date:

1. Formulați problema (întrebarea) cercetării
2. Formulați ipoteza cercetării (H1)
3. Formulați ipoteza de nul (H0)
4. Aflați ***t*** critic pentru α=0,05; bilateral
5. Calculați testul ***t*** pentru diferența dintre cele două eşantioane
6. Formulați şi motivați decizia statistică
7. Formulați concluzia cercetării

Verificați răspunsurile corecte la pagina 75

Analiza de varianță (mai mult de două eşantioane independente)

În situația în care am comparat performanța la țintă a celor două grupe de sportivi (practicanți şi nepracticanți de training autogen), testul ***t*** a rezolvat problema semnificației diferenței dintre două medii. În practica de cercetare ne putem întâlni însă cu situații în care avem de comparat trei sau mai multe medii. De exemplu, atunci când am efectuat un test de cunoştințe de statistică şi dorim să ştim dacă diferențele constatate între cele 5 grupe ale unui an de studiu diferă semnificativ. Performanța la nivelul fiecărei grupe este dată de media răspunsurilor corecte realizate de studenți. La prima vedere, am putea fi tentați să rezolvăm problema prin compararea repetată a mediei grupelor, două câte două. Din păcate, există cel puțin trei argumente pentru care această opțiune nu este de dorit a fi urmată:

- În primul rând, volumul calculelor ar urma sa fie destul de mare, şi ar creşte şi mai mult dacă numărul categoriilor variabilei independente ar fi din ce în ce mai mare.
- În al doilea rând, problema cercetării vizează relația dintre variabila dependentă (în exemplul de mai sus, performanța la statistică) şi variabila independentă, exprimată prin ansamblul tuturor categoriilor sale (grupele de studiu). Ar fi bine să putem utiliza un singur test şi nu mai multe, pentru a afla răspunsul la problema noastră.
- În fine, argumentul esențial este acela că, prin efectuarea repetată a testului ***t*** cu fiecare decizie statistică acumulăm o cantitate de eroare de tip I de 0.05 care se cumulează cu fiecare pereche comparată, ceea ce duce la depăşirea nivelului admis de convenția ştiințifică. Să presupunem că dorim să testăm ipoteza unei relații dintre nivelul anxietății şi intensitatea fumatului, evaluată în trei categorii: 1-10 țigări zilnic; 11-20 țigări zilnic şi 21-30 țigări zilnic. În acest caz, avem trei categorii ale căror medii ar trebui comparate două câte două. Dar, în acest fel, prin efectuarea repetată a testului t pentru eşantioane independente, s-ar cumula o cantitate totală de eroare de tip I de 0.15 adică 0.05+0.05+0.05.

Pentru a elimina aceste neajunsuri, şi mai ales pe ultimul dintre ele, se utilizează o procedură statistică numită analiza de varianță (cunoscută sub acronimul ANOVA, de la „***AN****alysis* ***O****f* ***VA****riance*", în engleză). În mod uzual, analiza de varianță este inclusă într-o categorie aparte de teste statistice. Motivul pentru care o introducem aici, imediat după testul ***t*** pentru eşantioane independente, este acela că, în esență,

Copyright © DEPARTAMENT ID 2008

ANOVA nu este altceva decât o extensie a logicii testului ***t*** pentru situațiile în care se dorește compararea a mai mult de două medii independente. Dar, dacă problema este similară, soluția este, așa cum vom vedea, diferită.

Există mai multe tipuri de ANOVA, două fiind mai frecvent folosite:

- **ANOVA unifactorială**, care se aplică atunci când avem o variabilă dependentă măsurată pe o scală de interval/raport măsurată pentru trei sau mai multe valori ale unei variabile independente categoriale. În contextul ANOVA, variabila independentă este denumită „factor", iar valorile pe care acesta le ia se numesc „niveluri". Din acest motiv, modelul de analiză de varianță cu o singura variabilă independentă se numește „ANOVA unifactorială", „ANOVA simplă" sau, cel mai frecvent, „ANOVA cu o singură cale" (*One-way ANOVA*).
    - Exemple:
        - Nivelul anxietății în raport cu trei categorii de fumători („1-10 țigări zilnic", „11-20 țigări" și „21-30 țigări").
        - Timpul de răspuns la un strigăt de ajutor, în funcție de natura vocii persoanelor care solicită ajutorul (copil, femeie, bărbat).
        - Scorul la un test de cunoștințe statistice ale studenților de la psihologie, în funcție de tipul de liceu absolvit (real, umanist, agricol, artistic).

- **ANOVA multifactorială**, care se aplică atunci când avem o singură variabilă dependentă (la fel ca în cazul ANOVA unifactorială) dar două sau mai multe variabile independente, fiecare cu două sau mai multe valori, măsurate pe o scală categorială (nominală sau ordinală).
    - **Exemple**
        - Nivelul anxietății în raport cu intensitatea fumatului („1-10 țigări zilnic", „11-20 țigări" și „21-30 țigări"), și cu genul (masculin, feminin). În acest caz, problema cercetării este dacă intensitatea fumatului și caracteristica de gen au, împreună, o relație cu nivelul anxietății.
        - Timpul de răspuns la un strigăt de ajutor în funcție de natura vocii care solicită ajutorul (copil, femeie, bărbat) și de genul (masculin, feminin) al persoanei care trebuie să răspundă la solicitarea de ajutor.
        - Scorul la un test de cunoștințe statistice ale studenților de la psihologie, în funcție de tipul de liceu absolvit (real, umanist, agricol, artistic) și de genul (masculin, feminin) al studenților.

Ne vom limita aici doar la prezentarea analizei de varianță unifactoriale, urmând să revenim cu alt prilej asupra altor variante de ANOVA.

Cadrul conceptual pentru analiza de varianță unifactorială

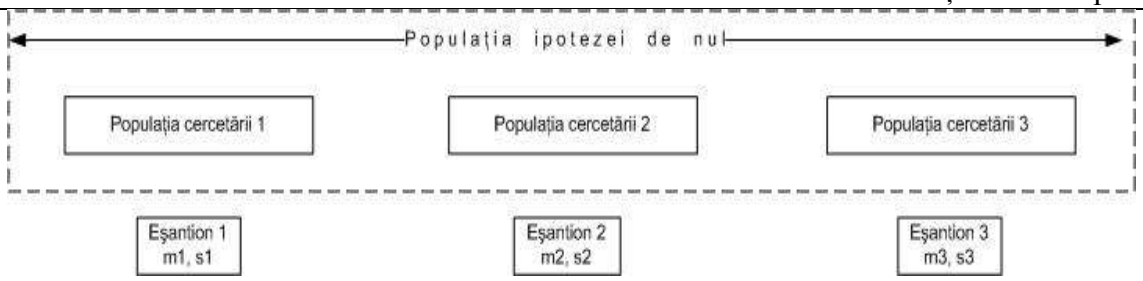
Să ne imaginăm o cercetare a cărei ipoteză este că relația dintre performanța sportivilor în tragerea la țintă și trei metode de antrenament (să le denumim metoda 1, metoda 2 și metoda 3).

În esență, ANOVA este o procedură de comparare a mediilor eșantioanelor. Specificul ei constă în faptul că în locul diferenței directe dintre medii se utilizează dispersia lor, gradul de împrăștiere. Procedura se bazează pe următorul demers logic: Ipoteza cercetării sugerează că performanța sportivilor antrenați cu fiecare dintre cele trei metode de antrenament face parte dintr-o populație distinctă, căreia îi corespunde un nivel specific de performanță (adică o medie caracteristică, diferită de a celorlalte două populații). Prin opoziție, ipoteza de nul ne obligă să presupunem că cele trei eșantioane[10] (modele de antrenament) pe care vrem să le comparăm, provin dintr-o populație unică de valori ale performanței, iar diferențele dintre mediile lor nu reprezintă decât expresia variației firești a distribuției de eșantionare.

În imaginea de mai jos populațiile cercetării (Pc1, Pc2, Pc3) sunt exprimate cu linie continuă, iar populație de nul cu linie discontinuă.

[10] Pentru simplificare, în continuare ne vom referi numai la trei eșantioane, dar se va înțelege „trei sau mai multe"

Copyright © DEPARTAMENT ID 2008



Chiar dacă absența unei legături între metoda de antrenament şi intensitatea nivelul performanței (ipoteză de nul) este adevărată, cele trei grupuri (eşantioane) nu trebuie să aibă în mod necesar aceeaşi medie. Ele pot avea medii diferite care să rezulte ca expresie a variației aleatoare de eşantionare ($m1 \neq m2 \neq m3$) şi, de asemenea, împrăştieri (dispersii) diferite ($s1 \neq s2 \neq s3$). Să ne gândim la cele trei medii pe care vrem să le comparăm, ca la o distribuție de sine stătătoare de trei valori (sau mai multe, pentru cazul în care variabila independentă are mai multe categorii). Cu cât ele sunt mai diferite una de alta, cu atât distribuția lor are o împrăştiere (varianță) mai mare. Este evident faptul că dacă eşantioanele ar aparține populației de nul, diferența mediilor (exprimată prin dispersia lor) ar fi mai mică decât în cazul în care acestea ar proveni din populații distincte (corespunzător ipotezei cercetării).

Mai departe, se pune următoarea problemă: cât de diferite (împrăştiate) trebuie să fie mediile celor trei eşantioane, luate ca distribuție de sine stătătoare de trei valori, pentru ca să putem concluziona că ele nu provin din populația de nul (dreptunghiul punctat), ci din trei populații diferite, corespunzătoare eşantioanelor de cercetare (Pc1, Pc2, Pc3)?

Pentru a răspunde la această întrebare este necesar:

a) Să calculăm dispersia valorilor individuale la nivelul populației de nul, care se bazează pe valorile performanței tuturor valorilor măsurate, indiferent de metoda de antrenament;
b) Să calculăm dispersia mediilor anxietății grupurilor cercetării (considerate ca eşantioane separate);
c) Să facem raportul dintre aceste două valori. Obținerea unei valori mai ridicate a acestui raport ar exprima apartenența fiecăreia din cele trei medii la o populație distinctă, în timp ce obținerea unei valori mai scăzute ar sugera proveniența mediilor dintr-o populație unică (de nul). Decizia statistică cu privire la mărimea raportului şi, implicit, cu privire la semnificația diferențelor dintre mediile comparate, se face prin raportarea valorii raportului la o distribuție teoretică adecvată, alta decât distribuția normală, aşa cum vom vedea mai departe.

În continuare ne vom concentra asupra fundamentării modului de calcul pentru cei doi termeni ai raportului. Calcularea exactă a dispersiei populației de nul este imposibilă, deoarece nu avem acces la toate valorile acesteia, dar poate fi estimată prin calcularea mediei dispersiei grupurilor de cercetare. Valoarea astfel obținută se numeşte „dispersia intragrup" şi reprezintă estimarea împrăştierii valorilor măsurate la nivelul populației de nul.

La rândul ei, dispersia mediilor grupurilor de cercetare, calculată după metoda cunoscută de calcul a dispersiei, formează ceea ce se numeşte „dispersia intergrup". Valoarea astfel obținută evidențiază cât de diferite (împrăştiate) sunt mediile eşantioanelor care fac obiectul comparației.

Raportul dintre „dispersia intergrup" şi „dispersia intragrup" se numeşte *raport F* şi ne dă valoarea testului ANOVA unifactorial. Cu cât acest raport este mai mare, cu atât împrăştierea mediilor grupurilor comparate este mai mare şi, implicit, diferența lor poate fi una semnificativă, îndepărtată de o variație pur întâmplătoare.

Imaginile de mai jos dau o expresie grafică acestui raționament:

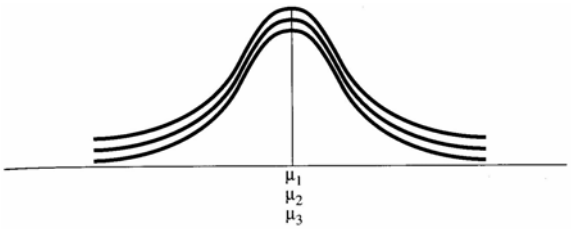


Figura a reprezintă grafic ipoteza de nul: presupunem că cele trei grupuri provin din aceeaşi populație. Ca urmare, cele trei medii sunt egale ($\mu_1=\mu_2=\mu_3$), iar distribuțiile sunt suprapuse.

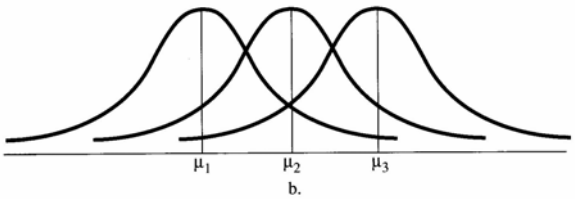


Figura b reprezintă grafic ipoteza cercetării: cele trei grupuri sunt diferite, provenind din populații distincte ($\mu_1 \neq \mu_2 \neq \mu_3$).

Copyright © DEPARTAMENT ID 2008

Dacă distanța (împrăştierea) dintre mediile eşantioanelor depăşeşte o anumită valoare, atunci putem concluziona că nu avem o singură populație (ipoteza de nul), ci mai multe, mediile grupurilor provenind din populații cu medii distincte (cf. ipotezei cercetării). Dacă, dimpotrivă, mediile eşantioanelor comparate sunt apropiate, atunci vom concluziona că ele nu provin din populații diferite, ci dintr-una singură (cf. ipotezei de nul).

Fundamentarea procedurii de calcul ANOVA[11]

Esența procedurii de calcul pentru ANOVA se bazează pe o dublă estimare a dispersiei:

(a) Estimarea dispersiei populației de nul pe baza mediei dispersiei grupurilor (varianța intragrup)

Atâta timp cât nu cunoaştem dispersia populației (σ^2) din care ar putea proveni grupurile, trebuie să o estimăm prin dispersiile celor trei grupuri (s_1^2, s_2^2, s_3^2).

Calculând media celor trei dispersii vom obține o valoare care estimează dispersia pentru cele trei grupuri luate împreună (indiferent de metoda de antrenament utilizată). Această valoare se consideră că estimează dispersia populației totale. Deoarece ea se calculează pe baza dispersiilor în interiorul grupurilor, este desemnată în mod uzual prin termenul de *intragrup* (sau, mai frecvent, prin forma engleză: *within-group*) şi se notează cu $s^2_{\text{intragrup}}$, fiind calculată cu una dintre formulele următoare:

Atunci când volumele eşantioanelor comparate sunt egale ($N_1=N_2=N_3$):

$$s^2_{\text{intragrup}} = \frac{s_1^2 + s_2^2 + s_3^2}{N_{grupuri}} \qquad \text{(formula 3.16)}$$

Atunci când grupurile comparate sunt de volum inegal:

$$s^2_{\text{intragrup}} = \frac{df_1}{df_{\text{intragrup}}} * s^2_1 + \frac{df_2}{df_{\text{intragrup}}} * s^2_2 + \frac{df_3}{df_{\text{intragrup}}} * s^2_3 \qquad \text{(formula 3.17)}$$

unde: $df_1=N_1-1$; $df_2=N_2-1$; $df_3=N_3-1$, iar $df_{\text{intragrup}}=N_{\text{subiecți}}-N_{\text{grupuri}}$

(b) Estimarea dispersiei populației de nul pe baza dispersiei mediilor grupurilor (varianța intergrup)

Mediile celor trei grupuri (eşantioane) sunt numere care pot fi analizate ca distribuție în sine, a căror dispersie (varianță) poate fi calculată, fiind o estimare a împrăştierii valorilor la nivelul populației. Din cauză că se bazează pe mediile grupurilor, aceasta se mai numeşte şi *varianță intergrupuri* (*between groups*, în limba engleză). Între variația acestor medii şi variația valorilor din grupurile analizate, luate împreună, există o legătură care poate fi exprimată pe baza formulei transformate a erorii standard, astfel:

$$\sigma^2_M = \frac{\sigma^2}{N} \quad \text{de unde se deduce} \quad \sigma^2 = N * \sigma^2_M \qquad \text{(formula 3.18)}$$

Vom putea utiliza dispersia mediilor celor trei eşantioane pentru a estima dispersia populației totale (vezi exemplul de mai jos). Aceasta se numeşte estimarea *varianței intergrupuri*, notată cu $s^2_{\text{intergrup}}$.

Dacă înlocuim în expresia de mai sus expresia de calcul a dispersiei (formula 3.17), obținem:

[11] Metoda de calcul pe care o vom prezenta aici (bazată pe dispersie) nu este singura posibilă. În multe manuale de statistică este utilizată metoda „sumei pătratelor", care se bazează pe scorurile brute, fără utilizarea parametrilor distribuției. Am preferat această metodă deoarece ni se pare mai intuitivă, pe de o parte, iar pe de altă parte, deoarece poate fi aplicată şi în cazul în care nu avem distribuția scorurilor brute, ci doar parametrii grupurilor comparate. Fiecare metodă are avantaje şi dezavantaje, dar în esență, ele conduc la acelaşi rezultat.

Copyright © DEPARTAMENT ID 2008

$$s^2_{\text{intergrup}} = \frac{\sum n_i * (m_i - M)^2}{df_{\text{intergrup}}} \quad \text{(formula 3.19)}$$

unde m_i este media performanței din fiecare grup, M este media celor trei grupuri luate împreună, iar n_i este numărul subiecților din fiecare grup, iar $df_{\text{intergrup}}$ se calculează ca numărul grupurilor-1.

Ca urmare, pentru o situație cu trei grupuri, formula desfăşurată se scrie astfel:

$$s^2_{\text{intergrup}} = \frac{n_1 * (m_1 - M)^2 + n_2 * (m_2 - M)^2 + n_3 * (m_3 - M)^2}{df_{\text{intergrup}}} \quad \text{(formula 3.19')}$$

unde: m_1, m_2, m_3 sunt mediile celor trei grupuri, n_1, n_2, n_3, sunt volumele celor trei eşantioane, iar celelalte valori sunt cele descrise pentru formula anterioară.

Pentru situația în care grupurile au un număr egal de subiecți, formula 3.19' devine:

$$s^2_{\text{intergrup}} = n * \frac{(m_1 - M)^2 + (m_2 - M)^2 + (m_3 - M)^2}{df_{\text{intergrup}}} \quad \text{(formula 3.19")}$$

unde n este numărul subiecților dintr-un grup.

Ambele tipuri de estimări sunt estimări independente ale varianței populației de nul. Însă, în timp ce varianța intragrup o estimează în mod direct (media varianțelor), varianța intergrup o măsoară indirect (varianța mediilor). Aceasta din urmă, varianța intergrup, reprezintă o estimare a varianței populației de nul numai dacă ipoteza de nul este adevărată. Dacă ipoteza de nul este falsă, ea reflectă de fapt măsura în care valorile variabilei independente (factorul) influențează mediile variabilei dependente. Pe această particularitate se bazează procedura analizei de varianță. Raportul dintre cele două estimări ($s^2_{\text{intergrup}}/s^2_{\text{intragrup}}$) va tinde să devină cu atât mai mare cu cât diferența dintre mediile grupurilor (tradusă prin dispersia mediilor) devine mai mare decât dispersia din interiorul grupurilor (tradusă prin media dispersiilor). Acest raport se numeşte „*raport Fisher*", după numele celui care a fundamentat acest tip de analiză[12], şi se scrie astfel:

$$F = \frac{s^2_{\text{intergrup}}}{s^2_{\text{intragrup}}} \quad \text{(formula 3.20)}$$

Interpretarea raportului F

Numitorul raportului F (dispersia intragrup) exprimă variabilitatea din interiorul grupurilor supuse comparației. Dacă analizăm sursele acestei variații, ea poate proveni din mai multe surse: diferențele individuale dintre subiecți, erorile de măsurare ale variabilei dependente, fluctuația condițiilor în care au fost efectuate măsurările. Neputând defini cu exactitate nici sursa şi nici contribuția fiecăreia, dispersia intragrup exprimă aşa numita „varianță neexplicată", definită generic şi ca „varianța erorii".

În conformitate cu ipoteza cercetării, grupurile de subiecți ar trebui să aibă scoruri diferite, fie pentru au fost supuse unui „tratament" diferit (în exemplul nostru prin cele trei metode de antrenament), fie ca urmare a faptului că fac parte din populații diferite. În acelaşi timp, subiecții din fiecare grup în parte ar trebui să aibă scoruri similare. Faptul că ele diferă totuşi, nu poate fi explicat prin efectul „tratamentului", motiv pentru care variația lor este definită drept o „varianță a erorii".

La rândul lui, variabilitatea numărătorului raportului F este rezultatul manipulării de către cercetător (atunci când operăm în context experimental), sau este rezultatul unor grupuri preexistente (atunci când efectuăm un studiu observațional). Şi valoarea acestuia este amplificată de varianța erorii. Aceasta deoarece, chiar şi în cazul în care „tratamentul" cu cele trei metode de antrenament ar fi total ineficient, şi toate populațiile ar avea medii identice, mediile grupurilor comparate ar diferi între ele, sub efectul unor surse diverse („erori"). Ca urmare, avem două surse de variabilitate la numărător şi numai una singură la numitor, fapt care poate fi sintetizat prin următoarea expresie:

[12] Sir Ronald Aylmer Fisher (1890-1962). Astronom de formație, interesat de teoria erorilor, s-a remarcat prin contribuțiile sale în teoria statisticii căreia, din anul 1922, i-a dat o nouă orientare.

Copyright © DEPARTAMENT ID 2008

$$F = \frac{\text{efectul tratamentului} + \text{varianța erorii intergrup}}{\text{varianța erorii intragrup}}$$

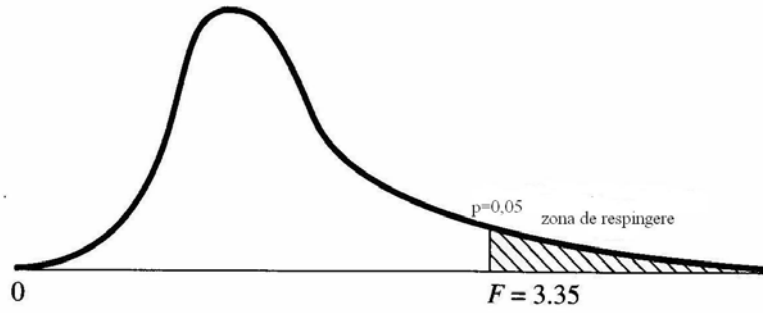
Atunci când ipoteza de nul este adevărată, efectul „tratamentului" se apropie de zero, iar raportul F este rezultatul variantei erorii. Dacă cele două varianțe ale erorii ar fi identice, F ar avea valoarea 1 dar, de fapt, cele două varianțe ale erorii pot avea valori diferite, ceea ce conduce la fluctuații ale lui F în jurul lui 1.

Atunci când efectul tratamentului nu este zero (ipoteza de nul este falsă), ne aşteptăm ca valoarea raportului F să fie mai mare decât 1. Însă pentru a respinge ipoteza de nul valoarea lui F trebuie să fie nu doar mai mare decât 1, ci mai mare decât un prag critic convențional asumat (alfa), astfel încât probabilitatea ca un rezultat similar să decurgă din întâmplare să fie mai mică sau cel mult egală cu alfa.

Distribuția Fisher

Valorile raportului F (sau testul F) se distribuie într-un mod particular, numit distribuția F sau distribuția Fisher. Ca şi distribuția normală, distribuția F este o familie de distribuții, având următoarele caracteristici:

1. asimetrie pozitivă (tendința valorilor de grupare spre partea stângă, cu valori mici);
2. poate lua valori oricât de mari;
3. valoarea minimă este 0, deoarece decurge din raportul a două dispersii, iar dispersiile nu pot fi niciodată negative[13].
4. forma distribuției variază în funcție de o pereche de grade de libertate formată din numărul grupelor (categoriile variabilei independente) şi numărul subiecților.



Imaginea de mai sus reprezintă curba F pentru 3 grupuri cu 30 de subiecți în total. Distribuția Fisher are forme distincte în funcție de numărul eşantioanelor comparate şi volumul acestora.

Calcularea gradelor de libertate

Ca şi în cazul distribuției ***t,*** distribuția F se prezintă sub o varietate de forme. Distribuția F rezultă dintr-un raport a două distribuții diferite ($s^2_{intergpup}$ şi $s^2_{intragrup}$), fiecare cu gradele ei de libertate. Ca urmare, îşi schimbă forma, în acelaşi timp în funcție de numărul grupurilor, şi de numărul subiecților din fiecare grup. În concluzie, vom avea două grade de libertate, unul pentru dispersia integrup şi altul pentru dispersia intragrup, calculate astfel:

$df_{intergrup}$=numărul grupurilor-1

$df_{intragrup}$=numărul cumulat al subiecților din toate grupurile-numărul grupurilor

EXEMPLU DE CALCUL

Problema cercetării:

Avem rezultatele la o şedință de tragere la țintă pentru trei grupuri de câte 6 sportivi, fiecare grup fiind antrenat cu o altă metodă, şi vrem să vedem dacă există o legătură între nivelul performanței şi metoda de antrenament.

[13] În practică, se poate ajunge în situația ca dispersia intragrup să rezulte a fi mai mică decât dispersia intergup şi, ca urmare, valoarea lui F să fie mai mică decât 0. Acest lucru este determinat de inegalitatea severă a dispersiilor între grupurile analizate.

Copyright © DEPARTAMENT ID 2008

Ipoteza cercetării:

„Performanța sportivă este în legătură cu metoda de antrenament utilizată.

Ipoteza de nul:

„*Nu există o legătură între performanța sportivă şi metoda de antrenament*."

Fixăm criteriile deciziei statistice:

Nivelul α=0.05

Stabilim F critic:

$df_{intergrup}$=3-1=2

$df_{intragrup}$=18-3=15

Citim F critic ($F_{(0.05, 2, 15)}$) din tabelul F pentru α=0.05:

F_{critic}=3.6823 (vezi tabelul anexat)

Notă privind utilizarea tabelei pentru distribuțiile F

Spre deosebire de tabelele distribuțiilor utilizate până acum, (***z*** şi ***t***), pentru interpretarea lui F avem mai multe tabele, calculate fiecare pentru un anume nivel al lui α. Mai întâi căutăm tabela pentru α dorit (să zicem, α=0.05). Apoi citim valoarea critică pentru F la intersecția dintre coloana care reprezintă numărul gradelor de libertate pentru numărul grupurilor (df_B) cu linia care reprezintă numărul gradelor de libertate pentru volumul total al subiecților (df_W). Dacă valoarea obținută prin calcul este mai mare sau egală decât cea tabelară, atunci putem lua decizia de respingere a ipotezei de nul.

O precizare importantă cu privire la ANOVA, ca test statistic, priveşte caracterul ei „unilateral" (*one-tailed*). Într-adevăr, spre deosebire de celelalte teste studiate până acum, ANOVA este interpretată într-o singură direcție şi anume, dacă mediile grupurilor diferă semnificativ între ele (au o variație mai mare decât cea normală pentru o distribuție aleatoare). Nu putem avea o valoare negativă pentru F şi, ca urmare, testul F este întotdeauna un test unilateral.

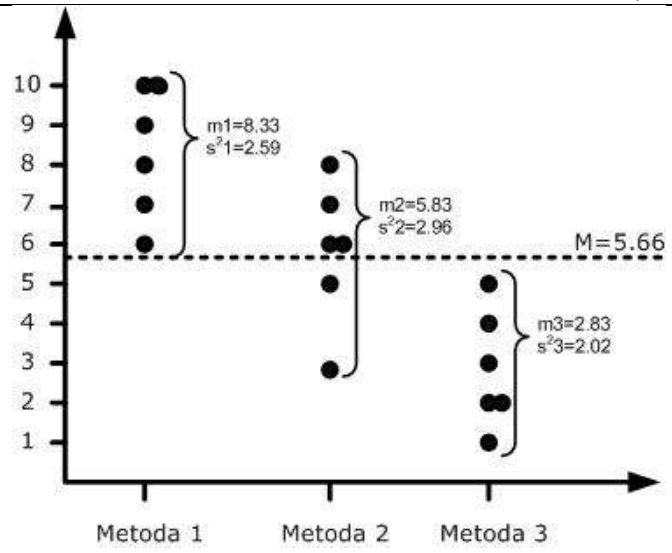
Calculăm F pe baza datelor centralizate în tabelul următor[14]**:**

	Metoda de antrenament					
	„metoda 1"		**„metoda 2"**		**„metoda 3"**	
	X1 (puncte)	**$(X1-m_1)^2$**	**X2 (puncte)**	**$(X2-m_2)^2$**	**X3 (puncte)**	**$(X3-m_3)^2$**
	10	2,79	3	8.00	4	1.36
	9	0,45	6	0.02	5	4.70
	10	2,79	6	0.02	2	0.68
	7	1,77	5	0.68	3	0.02
	8	0,11	8	4.70	2	0.02
	6	5,43	7	1.36	1	3.34
ΣX	50	13.33	35	14.78	17	10.14
N	6		6		6	
M	m_1=8.33		m_2=5.83		m_3=2.83	M=(m_1+m_2+m_3)/3=5.66
s^2	2.66		2.96		2.02	
(m-M)	2.67		0.17		-2.83	
$(m-M)^2$	7.12		0.02		8.00	$\Sigma(m-M)^2$=15.14

Distribuția valorilor celor trei grupuri poate fi ilustrată grafic astfel:

[14] Atenție, acest mod de prezentare a datelor serveşte calculării manuale a testului F. Într-o bază de date SPSS vom avea câte o înregistrare pentru fiecare subiect, cu două variabile, una pentru nivelul anxietății şi cealaltă pentru intensitatea fumatului, aceasta din urmă cu trei valori convenționale, să zicem 1, 2, 3 pentru fiecare nivel de intensitate a fumatului.

Copyright © DEPARTAMENT ID 2008



Recunoaştem în interiorul graficului parametrii fiecărui grup (m şi s^2) precum şi media „mare" (M), a valorilor individuale din toate grupurile, luate împreună.

Având calculați parametrii celor trei grupuri, putem trece la calcularea raportului F. Mai întâi calculăm numărătorul, adică dispersia mediilor celor trei grupuri. Dat fiind faptul că nu cunoaştem dispersia populației vom utiliza dispersia eşantioanelor, conform formulei 3.19" pentru grupuri egale.

Prin înlocuire cu valorile calculate în tabelul de mai sus, obținem:

$$s^2_{\text{intergrup}} = 6 * \frac{7.12 + 0.02 + 8.00}{2} = 6 * 7.57 = 45.42$$

Mai departe, calculăm numitorul raportului F (dispersia intragrup), prin înlocuirea valorilor calculate pentru dispersiile din interiorul celor trei grupuri luate separat, în formula 3.16:

$$s^2_{\text{intragrup}} = \frac{s_1^2 + s_2^2 + s_3^2}{N_{grupuri}} = \frac{2.66 + 2.96 + 2.02}{3} = 7.64$$

În acest caz $df_{\text{intragrup}}$=nr. grupurilor, pentru că N1=N2=N3
În final, calculăm raportul F:

$$F = \frac{s^2_{\text{intergrup}}}{s^2_{\text{intragrup}}} = \frac{45.42}{7.64} = 5.94$$

Valoarea astfel obținută o comparăm cu F critic găsit anterior în tabel. Constatăm că F calculat (5.94), este mai mare decât F critic (3.6823).

Decizia statistică:

Respingem ipoteza de nul şi acceptăm ipoteza cercetării: „*Nivelul performanței prezintă o variație în legătură cu metoda de antrenament utilizată*".

Mărimea efectului pentru testul F

La fel ca şi în cazul testelor statistice introduse anterior, valoarea testului F nu este informativă în sine. Mărimea lui F indică doar decât de câte ori este cuprinsă dispersia intragrup în dispersia intergrup. Pentru a decide dacă acest raport este „mare" sau „mic" trebuie să calculăm un indice al mărimii efectului. În cazul analizei de varianță sunt utilizați în mod obişnuit doi indici de mărime a efectului: *eta pătrat* (η^2) şi *omega pătrat* (ω^2). Spre deosebire de indicele d (Cohen), care este un indice al diferenței, *eta pătrat* şi *omega*

Copyright © DEPARTAMENT ID 2008

pătrat sunt indici ai asocierii[15] (B. Cohen, 2001), similari cu coeficientul de corelație, pe care îl vom analiza analiza în alt loc.

Vom prezenta aici doar indicele *eta pătrat*, dat fiind faptul că este accesibil cu metoda pe care am utilizat-o pentru calcularea lui F[16]. Formula de calcul pentru η^2 este următoarea:

$$\eta^2 = \frac{df_{\text{intergrup}} * F}{df_{\text{intergrup}} * \text{F} + \text{df}_{\text{intragrup}}} \quad \text{(formula 3.21)}$$

În esență, indicele *eta pătrat* descrie procentul din varianța (împrăştierea) variabilei dependente care este explicat de varianța variabilei independente.

Nu există o „grilă" unică de interpretare a indicelui *eta pătrat* dar, prin similitudine cu coeficientul de corelație, putem prelua sugestiile unor autori diferiți, ale căror opinii sunt, în linii mari, convergente. Redăm aici, pentru comparație, două variante de interpretare pentru *eta pătrat*:

Varianta de interpretare a lui Hopkins (2000):

0.9-1	Aproape perfect, descrie relația dintre două variabile practic indistincte
0.7-0.9	Foarte mare, foarte ridicat
0.5-0.7	Mare, ridicat, major
0.3-0.5	Moderat, mediu
0.1-0.3	Mic, minor
0.0-0.1	Foarte mic, neglijabil, nesubstanțial

Varianta de interpretare a lui Davis (citat de Kotrlik şi Williams, 2003)

0.70 →	asociere foarte puternică
0.50 – 0.69	asociere substanțială
0.30 – 0.49	asociere moderată
0.10 – 0.29	asociere scăzută
0.01 – 0.09	asociere neglijabilă

Vom observa că, în ambele variante, pentru a fi „important" indicele *eta pătrat* trebuie să atingă cel puțin valoare de 0.50, ceea ce înseamnă că 50% din varianță variabilei dependente este explicată de variabila independente.

Pentru datele exemplului nostru, indicele de mărime a efectului este:

$$\eta^2 = \frac{df_{\text{intergrup}} * F}{df_{\text{intergrup}} * \text{F} + \text{df}_{\text{intragrup}}} = \frac{2 * 5.94}{2 * 5.94 + 15} = 0.44$$

La rândul lui, Cohen (1988) a dezvoltat un indice de mărime a efectului (*f*) pentru ANOVA, care atenuează ceea ce se consideră a fi tendința de „supraestimare a mărimii efectului" de către indicele *eta pătrat*:

$$f = \sqrt{\frac{\eta^2}{1-\eta^2}} \quad \text{(formula 3.22)}$$

Pentru rezultatul din exemplul nostru:

$$f = \sqrt{\frac{0.44}{1-0.44}} = 0.88$$

[15] Fără a intra în amănunte, facem precizarea că indicii de mărime a efectului pot fi transformați cu ușurință unii într-alții, cu ajutorul unor formule de conversie.

[16] Programele statistice oferă, de regulă, posibilitatea de a calcula ambii indici ai puterii

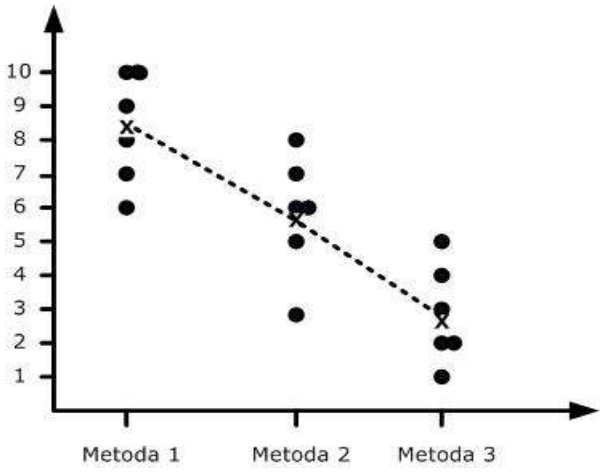
Copyright © DEPARTAMENT ID 2008

În conformitate cu recomandările lui Cohen, valorile lui *f* se interpretează astfel: efect mic=0.10; efect mediu=0.25; efect mare=0.40. Interpretarea mărimii efectului trebuie făcută cu precauție şi modestie (Runyon et. al, 1996). Un indice redus de mărime a efectului indică, desigur, o slabă intensitate a relației dintre variabila independentă şi variabila dependentă. Cu toate acestea, uneori, chiar şi o relație slabă între variabile poate fi importantă pentru cercetarea ştiințifică din ştiințele sociale şi umane. Comportamentul uman este supus unor surse extrem de complexe de determinări, fapt care face aproape imposibilă controlarea (eliminarea) unora dintre surse, pentru stabilirea exactă a efectului uneia anume. Acest lucru face inevitabilă prezența unei anumite cantități de erori de măsurare în toate cercetările psihologice. În aceste condiții, uneori, chiar şi un „efect mic" poate fi considerat un câştig important din punct de vedere ştiințific, chiar dacă este puțin relevant din punct de vedere practic. De exemplu, un rezultat semnificativ statistic, dar cu un indice scăzut de mărime a efectului, poate constitui punctul de plecare al unei noi cercetări, în care efectele colaterale ale unor variabile să fie mai bine controlate (eliminarea erorii), ceea ce poate conduce la evidențierea unei relații mai puternice între variabilele studiate.

Dacă privim cei doi indici ai mărimii efectului calculați pentru exemplul dat, putem aprecia că, în contextul datelor cercetării noastre, 44% din variația performanței de instruire este explicată de utilizarea metodelor de antrenament (ceea ce înseamnă, implicit, că un procent de 56% provine din alte surse). În conformitate cu recomandările de interpretare pentru *eta pătrat*, putem afirma că relația dintre metodele de antrenament utilizate şi performanță este „moderată" sau „medie". În acelaşi timp, indicele *f* al lui Cohen indică un nivel ridicat al mărimii efectului. Nu trebuie să privim aceste două aprecieri ale mărimii efectului ca fiind contradictoirii, ci ca pe două perspective asupra aceleiaşi realități.

Analiza „post-hoc"

Graficul alăturat prezintă variația mediilor performanței celor grupuri de sportivi. Aşa cum se observă, nivelul performanței are nivelul cel mai ridicat pentru prima metodă de antrenament (8.33), şi din ce în ce mai reduse la următoarele două (5.83; 2.83).



Testul ANOVA ne oferă o imagine „globală" a variației mediilor fără să ne spună nimic cu privire la „sursa" de proveniență acesteia, şi nici în ce măsură diferă mediile grupurilor luate dopuă cât două. În exemplul nostru valoarea obținută pentru F ar putea decurge doar prin „contribuția" unui singur grup (de ex., cei antrenați cu metoda 1), celelalte grupuri având o „contribuție" minoră sau inexistentă. Cercetătorul poate fi însă interesat care dintre grupuri diferă între ele, şi în ce sens.

Pentru a rezolva această problemă se efectuează aşa numitele comparații multiple, pe baza unor teste statistice denumite „*post-hoc*", pentru că, în mod normal, acestea se calculează după aplicarea procedurii ANOVA. Printre cele mai frecvent utilizate sunt testele: Scheffe, Tukey şi Bonferoni (desigur, se utilizează unul sau altul dintre ele, la alegere). Nu vom intra în detalii teoretice şi de calcul cu privire la aceste teste. Fiecare are avantajele şi dezavantajele sale. Important aici este să înțelegem că testele *post-hoc* se interpretează în mod similar testului ***t*** pentru diferența mediilor pentru eşantioane necorelate, calculate astfel încât să ia, atât cât se poate, măsuri de precauție împotriva excesului de eroare de tip I menționat anterior. Este important de reținut, de asemenea, faptul că analiza *post-hoc* este practicată, de regulă, numai dacă a fost obținut un rezultat semnificativ pentru testul F[17]. Aceasta înseamnă că analiza *post-hoc* nu poate fi utilizată ca substitut pentru testul ***t*** efectuat în mod repetat. Ca urmare, în practică, analiza de varianță va cuprinde două faze: prima, în care se decide asupra semnificației testului F, şi a doua, în cazul că acest raport este semnificativ, în care se analizează comparativ diferențele dintre categoriile analizate, pe baza unui test *post- hoc*.

În ce priveşte calcularea testelor *post-hoc* menționate mai sus, vom prezenta modul lor de calcul în secțiunea dedicată programului SPSS.

[17] Cu toate acestea, există autori care consideră că nimic nu ne împiedică să calculăm testele *post-hoc* chiar dacă testul F s-a finalizat cu admiterea ipotezei de nul.

Copyright © DEPARTAMENT ID 2008

Publicarea rezultatului testului F (ANOVA)

În raportul de publicare pentru ANOVA vor fi descrise grupurile (categoriile) comparate, mediile lor, valoarea testului F cu numărul gradelor de libertate şi pragul de semnificaţie al testului. La acestea se adaugă indicele de mărime a efectului. Într-o manieră narativă, rezultatul obținut pe exemplul de mai sus, poate fi prezentat astfel:

„A fost analizată performanța în tragerea la țintă a trei grupuri de sportivi, antrenați cu metode diferite. Mediile performanței pentru cele trei grupuri au fost 8.33, 5.83, respectiv 2.83. Analiza de varianță unifactorială a relevat o diferență semnificativă între aceste medii, F (2, 15)=6; p≤0.05. Mărimea efectului apreciată cu indicele eta pătrat indică un efect moderat (η^2=0.44), în timp ce indicele f al lui Cohen indică un efect mare (f=0.88)".

Atunci când vom calcula ANOVA cu ajutorul unui program care ne va oferi şi comparaţiile multiple între grupurile comparate (analiza *post-hoc*), la descrierea de mai sus vom adăuga şi comparaţiile grupurilor, două câte două, care exprimă diferenţele directe dintre grupurile supuse comparaţiei, explicând analitic sursele semnificaţiei raportului F global.

Avantajele ANOVA

Utilizarea ANOVA pentru testarea ipotezelor în cazul unui număr mai mare de grupuri (eşantioane) prezintă două avantaje. Primul, ţine de ceea ce am precizat deja, şi anume faptul că eliminăm riscul cumulării unei cantităţi prea mari de eroare de tip I, prin efectuarea repetată a testului ***t***. Al doilea, rezultă din faptul că avem posibilitatea să punem în evidenţă diferenţe semnificative între mediile mai multor grupuri, chiar şi atunci când nici una dintre ele nu diferă semnificativ una de cealaltă (testul ***t***).

Deşi, în mod normal, analiza de varianţă este utilizată doar în situaţia în care se doreşte testarea diferenţei dintre mediile a mai mult de două grupuri independente, ea dă rezultate echivalente şi în cazurile în care există numai două grupuri (singura diferenţă fiind valoarea calculată a testului, nu şi nivelul lui *p*). Utilizarea testului ***t*** pentru testarea diferenţei dintre două medii este, totuşi, o metodă mult mai directă, mai uşor de aplicat şi de înţeles, decât analiza de varianţă.

v. indep.	**v. dep.**
1	9
1	5
1	7
2	14
2	15
2	10

De exemplu, dacă luăm în considerare datele din tabelul alăturat, în care avem o variabilă dependentă distribuită pe două valori ale unei variabile independente, valoarea testului ***t*** este 3.13, iar valoarea testului F este 9.82 (ceea ce reprezintă pătratul valorii ***t***). În acelaşi timp, rezultatul la ambele teste este semnificativ pentru aceeaşi valoare a lui *p* (0.035).

Condiţii pentru utilizarea testului ANOVA

Utilizarea analizei de varianţă unifactoriale presupune îndeplinirea următoarelor condiţii:

- independenţa eşantioanelor (grupurilor supuse comparaţiei);
- normalitatea distribuţiei de eşantionare, în conformitate cu teorema limitei centrale;
- absenţa valorilor extreme (*outliers*);
- egalitatea varianţei grupurilor comparate (denumită „homoscedasticitate").

Atunci când una sau mai multe dintre aceste condiţii nu sunt întrunite, se poate adopta una dintre soluţiile următoare:

- renunţarea la ANOVA în favoarea unei prezentări descriptive (soluţie care ne lipseşte de posibilitatea unei concluzii testate statistic);
- transformarea variabilei dependente astfel încât să dobândească proprietăţile necesare (printre metodele uzuale, citãm aici doar logaritmarea sau extragerea radicalului din toate valorile variabilei dependente);

Copyright © DEPARTAMENT ID 2008

- transformarea variabilei pe o altă scală de măsurare şi aplicarea altui test statistic (de exemplu, prin transformarea pe o scală nominală, se poate aplica testul neparametric *chi-pătrat* sau, prin transformarea pe o scală ordinală, se poate aplica testul neparametric *Kruskal-Wallis*, ambele urmând a fi tratate mai departe).

Sarcina de lucru 2.5

Un psiholog doreşte să testeze ipoteza că preferința pentru o anumită varietate de bomboane este în legătură cu culoarea acestora (verde, roşu, galben). În acest scop alege 18 subiecți, pe care îi împarte în trei grupuri, fiecare grup primind câte un bol cu bomboane de o anumită culoare. Grupurile au de îndeplinit o sarcină plictisitoare timp de 30 de minute. După acest timp, psihologul numără câte bomboane a mâncat fiecare subiect din cele trei grupuri şi construieşte tabelul următor.

Verde	Roşu	Galben
2	3	2
1	4	0
1	5	2
0	6	1
3	4	3
2	6	1

1. Găsiți F critic pentru α=0.05
2. Calculați F
3. Care este decizia statistică în acest caz
4. Prezentați rezultatul în format APA

Verificați răspunsurile corecte.

Testul t pentru diferența dintre medii pentru eşantioane dependente

Testele de comparație prezentate până aici (***t*** pentru eşantioane independente şi ANOVA) au vizat situații în care mediile comparate aparțineau unor grupuri compuse din subiecți diferiți (motiv pentru care sunt denumite ca „independente", sau „necorelate"). Din cauză că acest model de cercetare presupune comparații între subiecți, el se mai numeşte şi model intersubiect (*between subject design*).

Un alt model uzual în cercetarea psihologică vizează comparația a două (sau mai multe) valori măsurate pe aceiaşi subiecți. Iată câteva ilustrări tipice:

a) Situația în care o anumită caracteristică psihologică se măsoară ***înaintea*** unei condiții şi apoi, ***după*** acțiunea acesteia. Exemple: (i) evaluarea nivelului anxietății înainte şi după un program de desensibilizare; (ii) evaluarea performanței cognitive a unui lot de subiecți, înainte şi după procedura de ascensiune simulată în camera barometrică la 5000m; (iii) evaluarea timpului de reacție înainte şi după ingerarea unei substanțe. Deoarece se bazează pe măsurări repetate ale unei variabile pe aceiaşi subiecți, acest model de cercetare este cunoscut ca „modelul măsurărilor repetate" (*repeated-measures design*).

b) Situația în care cercetătorul utilizează două condiții de investigare, dar plasează aceiaşi subiecți în ambele condiții. De exemplu, într-un studiu asupra efectelor unui anumit tip de stimulare, se pot măsura undele cerebrale, simultan în cele două emisfere cerebrale. Fiind vorba despre măsurarea unor variabile care sunt evaluate concomitent, la aceiaşi subiecți, acesta este un model „intrasubiect" (*within-subjects design*).

Copyright © DEPARTAMENT ID 2008

c) Cazul în care natura situației experimentale nu permite utilizarea aceloraşi subiecți pentru cele două măsurări, de exemplu, în contextul unei intervenții terapeutice care are un efect pe termen foarte lung. În acest caz este se poate găsi pentru fiecare subiect corespunzător condiției inițiale un subiect „similar", corespunzător condiției finale, constituind astfel „perechi de subiecți" aparținând fiecare unui grup distinct, între care se poate face o comparație directă. Ca urmare, deşi diferiți, vom trata cei doi subiecți din pereche ca şi cum ar fi aceeaşi persoană. Sau, într-un alt context, putem compara subiecți care sunt într-un anumit tip de relație, interesându-ne diferența dintre ei sub o anumită caracteristică. De exemplu, ne poate interesa daca între nivelul de inteligență dintre băieții şi fetele care formează cupluri de prieteni există o anumită diferență. În acest caz, deşi avem două eşantioane distincte, fiecărui subiect din eşantionul de băieți îi corespunde un subiect din eşantionul de fete, constituirea celor două eşantioane făcându-se pe baza relației de prietenie dintre ei. În aceeaşi categorie se află comparațiile între perechi de gemeni, sau cele dintre soți. În astfel de cazuri, avem de a face cu aşa numitul model al "eşantioanelor perechi" (*matched pairs design*).

Indiferent de tipul lor, toate modele prezentate mai sus au un obiectiv similar, acela de a pune în evidență în ce măsură o anumită condiție (variabila independentă) corespunde unei modificări la nivelul unei caracteristici psihologice oarecare (variabila dependentă). Vom observa că, în toate exemplele evocate, variabila independentă este una de tip nominal, dihotomic (înainte/după; semestru/sesiune; grup de cercetare/grup de control; băiat/fată; soț/soție, etc.), în timp ce variabila dependentă se măsoară pe o scală cantitativă, de interval sau de raport. De asemenea, trebuie să consemnăm faptul că în ambele situații se utilizează măsurători de acelaşi fel, cu acelaşi instrument, care produce valori exprimate în aceeaşi unitate de măsură, între care se poate efectua un calcul direct al diferenței.

Pentru descrierea testului statistic adecvat acestor cazuri să ne imaginăm următoarea situație generică de cercetare: Un grup de pacienți cu tulburări de tip anxios sunt incluşi într-un program de psihoterapie, având drept scop ameliorarea nivelului anxietății. Înainte de începerea programului a fost aplicată o scală de evaluare a anxietății. Acelaşi instrument a fost aplicat din nou, după parcurgerea programului de terapie.

Aici s-ar putea pune întrebarea de ce nu considerăm valorile rezultate din cele două măsurători ca fiind independente, urmând să utilizăm testul ***t*** pentru acest tip de date? Există mai multe argumente în favoarea respingerii acestei variante simplificatoare:

a) Utilizarea valorilor perechi oferă informații mai bogate despre situația de cercetare. În modele de cercetare de tip înainte/după ea capătă chiar valențe de experiment.
b) Testul ***t*** pentru eşantioane independente surprinde variabilitatea dintre subiecți, în timp ce testul ***t*** pentru eşantioane dependente (măsurări repetate) se bazează pe variabilitatea „intra-subiect", aceea care provine din diferența valorilor de la o măsurare la alta, la nivelul fiecărui subiect în parte.
c) Dacă există o diferență reală între subiecți, atunci testul diferenței dintre valorile perechi are mai multe şanse să o surprindă decât cel pentru valori independente (puterea unui model de cercetare *intra-subiect* este mai mare decât în modelul *inter-subiecți*).

Revenind la tema de cercetare pe care am enunțat-o mai sus, deşi avem aceiaşi subiecți, şi în primul şi în al doilea caz, ne vom raporta la aceasta situație ca şi cum ar fi două eşantioane. Unul, cel al subiecților care „nu au urmat încă" un program de terapie, iar celalalt, al subiecților care „au urmat" un astfel de program. Datorită faptului că cele două eşantioane sunt formate din aceiaşi subiecți, ele se numesc „dependente" sau „corelate".

În acest tip de studiu, obiectivul testului statistic este acela de a pune în evidență semnificația diferenței dintre mediile anxietății în cele două momente. Cea mai simplă procedură de calcul este metoda diferenței directe. Pentru aceasta, calculăm diferențele fiecărei perechi de valori din cele două distribuții (X2-X1), obținând astfel o distribuție a diferențelor, pe care o vom nota cu **D**.

Logica ipotezei de nul

Dacă programul de terapie ar fi total ineficient, trebuie să presupunem că diferențele pozitive le-ar echilibra pe cele negative ceea ce, la un număr mare de eşantioane ipotetice (formate din acelați număr de subiecți), am obține o medie a diferențelor egală cu 0. Ca urmare, ipoteza statistică presupune că media diferențelor la nivelul populației de nul este 0. Aceasta înseamnă că testul ***t*** trebuie să demonstreze că media diferențelor măsurate pe eşantionul cercetării este suficient de departe de 0, pentru a respinge ipoteza de nul şi a accepta ipoteza cercetării. De aici rezultă că putem reduce metoda de calcul la formula testului ***t*** pentru un singur eşantion, pornind de la formula cunoscută a testului ***t,***

Copyright © DEPARTAMENT ID 2008

$$t = \frac{m - \mu}{s_m}$$

Numitorul, eroarea standard a diferenței dintre medii, se calculează cu formula:

$$s_{eD} = \frac{s_D}{\sqrt{N}} \text{ (formula 3.23)}$$

Ca urmare, formula pentru testul t al diferenței dintre medii dependente este:

$$t = \frac{m_D - \mu_D}{s_{eD}} \quad \text{(formula 3.24)}$$

unde m_D este media distribuției D (a diferențelor dintre cele două măsurări), μ_D este media populației de nul a diferențelor dintre eşantioane de acelaşi fel, iar s_{eD} este eroarea standard a distribuției D (împrăştierea distribuției D).

Exemplu analitic de calcul

Problema cercetării: Se poate obține o reducere a reacțiilor anxioase prin aplicarea unei anumite proceduri de psihoterapie?

Ipoteza cercetării (H_1):

Pentru test bilateral → Programul de psihoterapie are un efect asupra anxietății.

Pentru test unilateral → Programul de psihoterapie reduce intensitatea reacțiilor de tip anxios.

Ipoteza de nul (H_0):

Pentru test bilateral → Programul de psihoterapie nu are nici un efect asupra anxietății.

Pentru test unilateral → Programul de psihoterapie nu reduce nivelul anxietății.

Populațiile cercetării:

Populația 1 → Subiecții cu anxietate ridicată care nu au urmat un program de terapie

Populația 2 → Subiecții cu anxietate ridicată care au urmat un program de terapie

Ipoteza cercetării afirmă că ele sunt diferite (m_1-$m_2 \neq 0$), în timp ce ipoteza de nul afirmă că ele sunt identice (m_1-m_2=0).

Eşantion: Un singur grup de subiecți cu probleme anxioase (N=8) al cărui nivel de anxietate este evaluat înainte şi după programul de terapie.

Criteriile deciziei statistice

Alegem modul de testare a ipotezei, bilateral.

Fixăm, convențional, nivelul α=0.01. Să spunem că preferăm acest nivel deoarece costurile de implementare a programului sunt destul de mari, iar pacienții trebuie convinşi că merită timpul şi banii[18].

Căutăm ***t*** critic pentru α=0.01 în tabelul distribuției ***t*** pentru 7 grade de libertate (N-1). Tabelul ne dă valorile pentru un test unilateral (în dreapta curbei). Pentru testul bilateral trebuie mai întâi să înjumătățim valoarea aleasă pentru α (0.01/2=0.005). În continuare, căutăm valoare aflată la intersecția coloanei gradelor de libertate (7) cu coloana lui α=0.005 şi citim ***t*** critic= -3.49. Îi atribuim semnul minus, deoarece ne aşteptăm ca nivelul anxietății să scadă după aplicarea programului de terapie.

Datele cercetării:

Înainte de program (X1)	**După program (X2)**	**D (X2-X1)**	**D-m_D**	**(D-m_D)^2**
6	6	0.00	0.50	0.25
8	7	-1.00	-0.50	0.25

[18] Am optat pentru alfa=0.01 doar pentru a varia exemplele de calcul, dar in practică se utilizează în mod obişnuit alfa=0.05.

Copyright © DEPARTAMENT ID 2008

	Înainte de program (X1)	După program (X2)	D (X2-X1)	D-m_D	$(D-m_D)^2$
	10	11	1.00	1.50	2.25
	9	8	-1.00	-0.50	0.25
	5	5	0.00	0.50	0.25
	6	5	-1.00	-0.50	0.25
	11	10	-1.00	-0.50	0.25
	5	4	-1.00	-0.50	0.25
ΣX	60	56	-4		$\Sigma(D-m_D)^2=4$
N	8	8	8		
$m=\frac{\sum X}{N}$	7.50	7.00	m_D=-0,5		
$s_D=\sqrt{\frac{(D-m_D)^2}{N-1}}$					$s_D=\sqrt{\frac{4}{7}}=0.75$

Notă: În principiu, sub aspectul procedurii statistice, nu prezintă nici o importanță dacă utilizăm diferența X1-X2 sau X2-X1. Ordinea depinde de ceea ce doreşte să scoată în evidență cercetătorul. Important este ca, în final, să interpreteze corect rezultatul obținut, în funcție de semnul diferenței şi semnificația concretă a acestuia.

Introducem valorile în formula 3.24 şi obținem:

$$t=\frac{-0,5}{0,75/\sqrt{8}}=-2.08$$

Raționamentul decizional

- Comparăm ***t*** calculat cu t critic pentru α=0.01 bilateral: -2,08 < -3.49
- Decizia statistica: „*acceptăm ipoteza de nul*”. Probabilitatea de a se obține un nivel al anxietății mai redus doar ca urmare a jocului hazardului, este mai mare decât nivelul alfa pe care ni l-am impus drept criteriu de decizie (adică mai mic de 1%).
- Decizia cercetării: „datele nu sprijină ipoteza cercetării”. Ca urmare, nu putem accepta că efectul obținut se datorează programului de terapie. Programul de terapie nu reduce în mod semnificativ nivelul anxietății.

Mărimea efectului

Indicele de mărime a efectului (*d - Cohen*) pentru diferența dintre medii dependente se calculează cu formula lui Cohen:

$$d=\frac{m_2-m_1}{s_D} \text{ (formula 3.25)}$$

Interpretarea indicelui *d* se face în conformitate cu recomandările lui Cohen, astfel: 0.20, efect mic; 0.50, efect mediu, 0.80, efect mare.

Pentru exemplul nostru, indicele de mprime a efectului este:

$$d=\frac{m_2-m_1}{s_D}=\frac{7-7.50}{0.75}=-0.66$$

Valoarea obținută indică o diferență „medie-mare” sau „relativ importantă” între mediile comparate (semnul lui *d* nu are relevanță). Aşa cum se vede, este posibil să obținem un indice al mărimii efectului

Copyright © DEPARTAMENT ID 2008

„mediu spre ridicat" în condițiile unui rezultat nesemnificativ statistic. Acest lucru trebuie să ne atragă odată în plus atenția asupra faptului că cele două proceduri (testul statistic şi mărimea efectului) vizează aspecte diferite. Pentru exemplul nostru, vom concluziona că efectul terapiei este relativ important, dar nu are o putere suficientă penmtru a atinge pragul de semnificație pe un lot de numai 8 subiecți. Este mai mult decât probabil că pe un eşantion mai mare rezultatul ar atinge şi pragul de semnificație statistică.

Limitele de încredere pentru diferența dintre medii

La fel ca şi în cazul testului ***t*** pentru eşantioane independente, se pune problema generalizării rezultatului la nivelul populației, cu alte cuvinte, care este intervalul în care ne putem aştepta să se afle diferența dintre medii, pentru variabilele studiate. Pentru o estimare cu o precizie de 99%, conform cu nivelul alfa ales, limitele critice pentru diferența dintre medii sunt cele care corespund valorilor lui p=0,005, de o parte şi de alta a curbei ***t*** (±3.4998). Formula de calcul pentru intervalul de încredere derivă, şi în acest caz, din formula 3.24:

$$t = \frac{m_D - \mu_D}{s_{eD}}$$

de unde rezultă formula pentru calculul limitelor de încredere ale mediei diferenței:

$$\mu_D = m_D \pm t_{crit} * s_{eD} \quad \text{(formula 3.26)}$$

În condițiile studiului nostru, decizia statistică de acceptare a ipotezei de nul a infirmat ipoteza cercetării dar analiza intervalului de încredere poate ajuta la înțelegerea mai bună a situației. Înlocuind valorile corespunzătoare studiului nostru, obținem următoarele limite de încredere:

limita inferioară: μ_D = -0.5-(-3.4998)*0.26= +0.40

limita superioară μ_D = -0.5+(-3.4998)*0.26=-1.4

Rezultatul arată că media diferenței la nivelul populației se află, cu o probabilitate de 0.99 (sau 99%), între o limită inferioară=+0.40 şi o alta superioară=-1.40. În acest caz, „inferior" se referă la o valoare plasată în jumătatea stângă a curbei t, unde valori inferioare sunt cele care se apropie de 0, care este media diferenței de nul. Aşa cum se constată, intervalul de încredere cuprinde şi valoarea 0, care exprimă ipoteza de nul (diferență nulă). Acest lucru este concordant cu decizia statistică, în urma căreia am admis ipoteza de nul şi am respins ipoteza cercetării. O privire mai atentă asupra datelor ar putea să îi arate cercetătorului că unul dintre subiecți a obținut un scor mai mare al anxietății după terapie decât înainte de terapie, fapt care este nefiresc şi ar trebui analizat. Acest caz se pare ca a fost decisiv în neatingerea pragului de semnificație. O reluare a procesului de diagnostic psihologic cu subiectul în cauză poate, eventual, conduce la concluzia că problemele lui sunt de altă natură (de ex., suferă de depresie şi nu de anxietate) şi că, în cazul său, terapia respectivă nu are nici un efect. Refacerea calculelor cu scoaterea din eşantionul de cercetare a acestui subiect (numai dacă acest lucru este bine motivat), va conduce, cu siguranță, la un interval mai restrâns de încredere pentru diferența dintre medii, ceea ce va însemna o precizie de estimare mai ridicată şi, implicit, poate, la atingerea pragului de semnificație.

Nu trebuie să omitem, de asemenea, faptul că în exemplul nostru este vorba de un eşantion foarte mic, iar eşantioanele mici conduc la valori ridicate ale erorii standard a mediei şi, prin aceasta, la intervale de încredere largi. În astfel de situații riscul erorii de tip II (imposibilitatea de a pune în evidență diferențe reale, rezultat fals negativ) este mai mare. Dar, atunci când obținem rezultate semnificative pe eşantioane mici, ele pot prezenta un nivel de încredere cu atât mai mare. În acelaşi timp, eşantioanele mici sunt instabile (în exemplul nostru, o singură diferență pozitivă poate schimba rezultatul cercetării), fapt care impune cel puțin replicarea cercetării, pentru mai multă siguranță.

Copyright © DEPARTAMENT ID 2008

Publicarea rezultatului

La publicare se vor menționa: volumul eşantionului, mediile variabilei dependente în raport cu valorile variabilei independente, valoarea testului t, pragul de semnificație, tipul de test (unilateral sau bilateral), mărimea efectului şi limitele de încredere ale diferenței. Având în vedere faptul că, uzual, testele statistice se efectuează bilateral, se poate menționa numai cazul în care testul este unilateral, eventual cu explicarea motivului pentru care a fost preferată această soluție.

Pentru exemplul de mai sus, o prezentare narativă a rezultatului ar putea arăta astfel:

„*Un eşantion de 8 subiecți cu probleme de anxietate au participat la un program de terapie anxiolitică. Nivelul anxietății (măsurat cu o scală specifică) a fost evaluat înainte şi după programul de terapie. S-a constatat o reducere a nivelului anxietății de la o medie de 7.50 la 7.0, după aplicarea terapiei. Diferența nu a atins pragul semnificației statistice t(7)=-2,08, p<0.01, pentru α=0.01 bilateral, cu limitele de încredere (99%) cuprinse între +0.40 şi -1.40. Indicele d (Cohen) al mărimii efectului (0.66) arată totuşi existența unei diferențe relativ importante între mediile celor două momente. Absența semnificației statistice se datorează, foarte probabil, volumului foarte redus al eşantionului şi existenței unui scor extrem al unuia dintre subiecți. În concluzie, rezultatele încurajează utilizarea în continuare a metodei terapeutice şi reevaluarea eficienței ei pe un eşantion mai mare.*"

Sarcina de lucru nr. 2.6

Un psiholog îşi propune să scoată în evidență efectul stresului temporal (criza de timp) asupra performanței de operare numerică. În acest scop, selectează un eşantion de subiecți cărora le cere să efectueze un test de calcule aritmetice în două condiții experimentale diferite: prima, în condiții de timp nelimitat, cu recomandarea de a lucra cât mai corect; a doua, în condiții de timp limitat, cu condiția de a lucra cât mai repede şi mai corect în acelaşi timp.

Rezultatele celor două reprize sunt cele din tabelul alăturat.

Fără criză de timp	Cu criză de timp
67	65
79	73
83	70
80	85
99	93
95	88
80	72
100	69

Să se rezolve următoarele sarcini:

1. Formularea ipotezei cercetării şi a ipotezei de nul
2. Stabilirea valorii t critic pentru **α=0,05** bilateral
3. Calcularea testului t
4. Decizia statistică
5. Decizia cercetării
6. Formularea concluziei în raportul de cercetare (format APA)

Verificați răspunsurile corecte.

Copyright © DEPARTAMENT ID 2008

Testarea asocierii dintre două variabile măsurate pe aceiaşi subiecți

Coeficientul de corelație liniară (Pearson)

Am utilizat testul *t* pentru eşantioane dependente pentru a evalua semnificația diferenței dintre două medii, rezultate în urma măsurării unei variabile cantitative, pe un eşantion de subiecți (sau de subiecți „pereche") aflat în două situații (condiții) diferite. Cele două condiții reprezintă valorile variabilei independente, iar cercetătorul este interesat să afle dacă există o diferență semnificativă între mediile rezultate în fiecare dintre cele două condiții. În concluzie, problema cercetării într-o situație de acest gen este axată pe ideea de „diferență între perechile de valori".

În practica cercetării există fie situații în care nu suntem interesați de „diferența dintre două medii" rezultate pe acelaşi grup de subiecți, ci de „gradul de asociere dintre două variabile măsurate pe acelaşi grup de subiecți. fiecare variabilă reprezentând altceva. În acest caz cercetătorul doreşte să afle dacă există o legătură între variația valorilor unei variabile în raport cu cealaltă variabilă.

Pentru a înțelege mai bine diferența dintre cele două abordări statistice, să ne uităm puțin în tabelele de date de mai jos:

	v1	v2
sub 1	7	4
sub 2	9	5
sub 3	8	6
sub 4	6	7
sub 5	5	3
media	7	5

(a) În cazul diferenței dintre medii, valorile celor două distribuții (v1 şi v2) pentru un eşantion de 5 subiecți sunt „condensate" prin mediile lor (7 şi 5), a căror diferență (7-5=2) este testată din punct de vedere al semnificației statistice.

	v1	v2
sub 1	7	4
sub 2	9	5
sub 3	8	6
sub 4	6	7
sub 5	5	3

(b) În cazul corelației dintre valorile celor două distribuții se urmăreşte punerea în evidență a modului în care se asociază valorile perechi, adică în ce măsură există o legătură între fiecare valoare şi perechea ei.

Într-o relație de asociere ambele variabile sunt dependente una de alta, iar valorile lor pot fi exprimate, fie cu aceeaşi unitate de măsură, fie cu unități de măsură diferite. Iată câteva exemple de acest gen:

- *există o legătură între numărul atitudinilor pozitive pe care le manifestă oamenii şi numărul atitudinilor pozitive pe care le primesc din partea celor din jur?*
- *există o legătură între timpul de reacție şi nivelul extraversiunii, ca trăsătură de personalitate?*
- *există o legătură între greutate şi înălțime?*
- *există o relație între frecvența pulsului şoferilor şi viteza cu care conduc maşina?*
- *există o relație între numărul orelor de studiu la statistică şi punctajul obținut la evaluări?*

În toate aceste situații avem câte două variabile, ambele fiind dependente una în raport cu alta, în sensul că este vizată existența unei concordanțe în variația reciprocă a valorilor celor două variabile, iar testul statistic utilizat se bazează pe calcularea unui „coeficient de corelație".

Înainte de a fi un concept statistic termenul de corelație este un cuvânt uzual în limbajul cotidian. În esență, el exprimă o legătură între anumite aspecte ale realității, aşa cum este ea reflectată în planul observației directe. De exemplu, o parcare plină cu maşini ne sugerează că magazinul alăturat este plin cu cumpărători, între numărul de maşini din parcare şi numărul de cumpărători existând o anumită „corelare". La nivel statistic, corelația exprimă o legătură cantitativă sistematică între valorile a două variabile perechi, măsurate pe subiecți aparținând aceluiaşi eşantion de cercetare. Coeficientul de corelație este doar una dintre procedurile statistice prin care se pune în evidență „corelarea" dintre variabile. În termeni generali, chiar şi testele *t*, prezentate anterior, pun în evidență (co)relația (legătura) dintre o variabilă dependentă şi valorile unei variabile independente.

Copyright © DEPARTAMENT ID 2008

Corelația liniară

Să presupunem că un grup de studenți au efectuat un test de inteligență bazat pe raționament abstract/figurativ şi un altul, bazat pe raționament verbal/logic. Dacă studenții care obțin valori mari la primul test tind să obțină valori mari şi la cel de-al doilea, avem ceea ce se numeşte o corelație pozitivă. Dacă, dimpotrivă, studenții care obțin valori mari la unul dintre teste tind să obțină valori mici la cel de-al doilea, atunci ne aflăm în fața unei corelații negative. Este evident că există şi posibilitatea ca valorile celor două variabile să evolueze absolut independent unele de celelalte, ceea ce indică absența oricărei corelații.

Precursorul teoretic al coeficientului de corelație este coeficientul de covarianță. El se defineşte ca sumă a produselor dintre valorile celor două variabile, raportată la numărul perechilor de valori din cele două distribuții:

$$\mathrm{cov}_{xy} = \frac{\sum x * y}{N} \qquad \text{(formula 3.27)}$$

unde x şi y sunt valorile perechi ale celor două variabile, iar N este volumul eşantionului.

Problema pe care o ridică coeficientul de covarianță este legată de unitățile de măsură. Formula poate fi aplicată numai dacă valorile perechi sunt exprimate în aceeaşi unitate de măsură, pentru ca produsul lor să aibă sens. Evident, nu am putea-o utiliza pentru a calcula coeficientul de covariație între înălțime şi greutate, de exemplu, deoarece este dificil să înțelegem rezultatului unui produs dintre unități de măsură diferite (greutate şi lungime). Soluția problemei constă în transformarea valorilor celor două variabile în scoruri standard, ceea ce produce un rezultat care nu mai are legătură cu unitatea de măsură. Intensitatea legăturii dintre valorile a două variabile se exprimă prin coeficientul de corelație liniară, notat cu simbolul ***r***. Introdus de Karl Pearson[19], acest coeficient mai este cunoscut şi sub numele de coeficientul de corelație Pearson, sau al „*moment-produsului*", după expresia uneia din formulele de calcul.

Formula de definiție a coeficientului de corelație este:

$$r = \frac{\sum z_x * z_y}{N} \qquad \text{(formula 3.28)}$$

unde z_x respectiv z_y sunt scorurile z ale variabilelor x şi y, iar N este volumul eşantionului.

Situația de maximă corelație posibilă între cele două distribuții este atunci când valorile lor sunt identice. Dacă ar fi aşa, atunci valorile z_x sunt egale cu valorile z_y, iar formula 3.28 ar putea scrisă ca:

$$r = \frac{\sum z_x^{\,2}}{N} \qquad \text{(formula 3.29)}$$

În continuare, dacă înlocuim în formula de mai sus expresia de calcul a lui z şi facem toate simplificările posibile, ajungem în final la formula deja cunoscută a dispersiei. În consecință, din faptul că dispersia unei distribuții z este întotdeauna egală cu +1, rezultă că valoarea maximă pe care o poate atinge coeficientul de corelație, în cazul unei corelații pozitive perfecte, este r=+1. Corespunzător, în cazul unei corelații negative perfecte, conform aceluiaşi raționament, rezultă că valoarea minimă posibilă a coeficientului de corelație este r= -1.

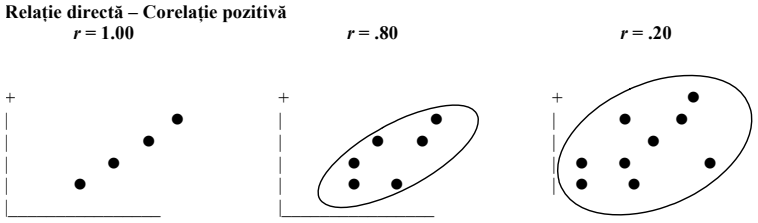
Reprezentarea grafică a corelației

Plasarea valorilor corelate pe un grafic, produce o imagine intuitivă a relației dintre valori. Acest tip de grafic se numeşte *scatterplot*.

În cazul unei corelații pozitive, reprezentările *scatterplot* arată astfel:

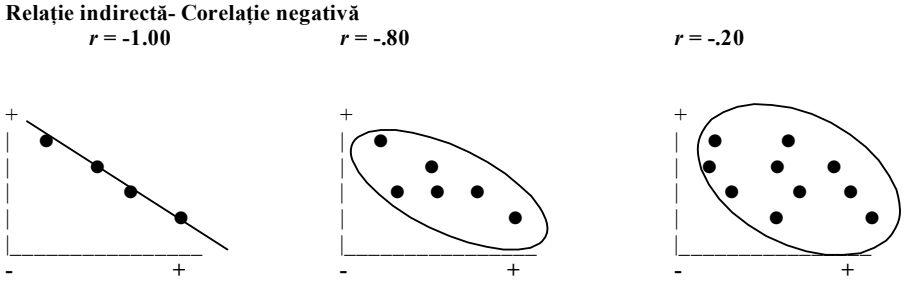
[19] Karl Pearson (1857-1936), matematician, filozof al ştiinței, biometrician şi statistician englez.

Copyright © DEPARTAMENT ID 2008



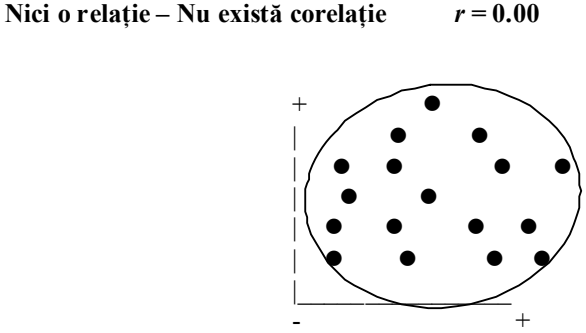
În cazul unei corelații pozitive valorilor mari ale unei variabile tind să le corespundă valori mari le celeilalte variabile. La limită, pentru o corelație pozitivă perfectă (r=+1) punctele de intersecție ale perechilor de valori se plasează pe o linie dreaptă, dinspre stânga jos spre dreapta sus, la un unghi de 45 de grade între cele două axe. Cu cât corelația este mai mică, cu atât norul de puncte este mai larg, dar forma elipsei indică relația pozitivă dintre cele două variabile.

În imaginea de mai jos avem reprezentări *scatterplot* caracteristice pentru trei corelații liniare negative.



În cazul corelației negative, tendința este aceea ca valorilor mari ale unei variabile să le corespundă valori mici ale celeilalte variabile. Ca urmare, atât linia corelației negative perfecte (r=-1), cât şi diagonala mare a elipsei norului de puncte al unei corelații negative imperfecte, se orientează din stânga sus spre dreapta jos a sistemului de coordonate.

În fine, atunci când corelația dintre cele două variabile este inexistentă, norul punctelor de intersecție are o formă circulară, care nu conturează nici o tendință (r=0).



Calcularea coeficientului de corelație liniară Pearson

De obicei, pentru a uşura calcularea manuală a coeficientului de corelație, mai ales atunci când avem date numeroase, sunt utilizate formule derivate din formula de definiție (formula 3.28), prin înlocuirea expresiilor pentru scorul z.

$$r = \frac{\sum z_x * z_y}{N} = \frac{\sum \left[(X - m_x)/s_x\right] * \left[(Y - m_y)/s_y\right]}{N}$$

Se deduce astfel o formulă care, deşi apare mai complicată, este mai uşor de pus în practică, deoarece se bazează pe valori care se obțin prin calcule mai simple:

Copyright © DEPARTAMENT ID 2008

de unde obținem: $r = \frac{\sum (X - m_x) * (Y - m_y)}{N * s_x * s_y}$ (formula 3.30)

unde:

- X şi Y reprezintă valorile individuale ale distribuțiilor X şi Y
- m_x şi m_y reprezintă mediile distribuțiilor X şi Y
- s_x şi s_y reprezintă abaterile standard ale distribuțiilor X şi Y
- N este volumul eşantionului

Formula 3.30 este doar una dintre variantele utilizate. Ea poate fi utilizată pentru calcule, la fel de bine ca şi formula 3.28, obținându-se rezultate identice. În general, pentru păstrarea acurateței rezultatului se recomandă păstrarea primelor patru zecimale ale fiecărei operații de calcul dar, pentru exemplele didactice, unde rezultatul nu are o finalitate reală, se poate lucra şi cu primele două zecimale. Oricum, în final, valoarea coeficientului *r* se raportează în mod obişnuit cu doar două zercimale.

Criteriile deciziei statistice

La fel ca în cazul celorlalte teste statistice, şi coeficientul *r* se raportează la o distribuție teoretică, care este una derivată din distribuția ***t***. Indiferent de cât de mare este *r* calculat, nu putem avea încredere în acesta atâta timp cât nu ştim în ce măsură este diferit de un *r* care ar rezulta prin jocul întâmplării. Pentru aceasta se utilizează distribuția ***t*** şi o formulă care derivă din testul ***t***.

Pentru uşurarea evaluării semnificației, a fost creat un tabel special cu praguri de semnificație ale coeficientului de corelație *r* care poate fi folosit fără a mai fi necesară utilizarea formulei (vezi tabelul semnificațiilor coeficientului de corelație din anexă). Practic, se caută în tabel care este nivelul lui *r* pentru numărul gradelor de libertate (df=N-2), şi un prag α ales în prealabil. Dacă valoarea calculată este cel puțin egală sau mai mare decât valoarea tabelară (critică) a lui *r*, atunci ipoteza de nul se respinge, coeficientul de corelație fiind considerat semnificativ.

Pentru exemplul nostru, pentru test unilateral, α=0.05 şi df=6 (8-2), citirea tabelului se face ca în figura alăturată.

df	Nivel de semnificaţie – test unilateral			
	.05	.025	.01	.005
	Nivel de semnificaţie – test bilateral			
	.10	.05	.02	.01
1	.988	.997	.9995	.9999
2	.900	.950	.980	.990
3	.805	.878	.934	.959
4	.729	.811	.882	.917
5	.669	.754	.833	.874
6 →	.622	.707	.789	.834

În condițiile precizate pentru cercetarea propusă ca exemplu, valoarea tabelară (critică) a lui *r* este 0.622. Dacă am fi preferat un test bilateral, pentru acelaşi nivel al lui alfa, valoarea *r* critic ar fi fost 0.707.

EXEMPLU DE CALCUL

Vom lua în considerare cazul aplicării celor două teste de raționament de tip diferit. În acest caz, ipoteza cercetării se exprimă în maniera: „*există o legătură (corelație) între cele două tipuri de raționament, cei care obțin rezultate bune la unul din teste, vor tinde sa obțină rezultate bune şi la celalalt”. D*esigur, ipoteza poate fi formulată şi corespunzător unei corelații negative, dacă avem motive să presupunem acest lucru.

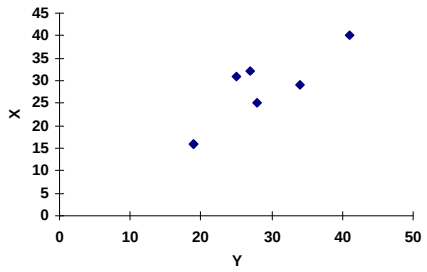
Copyright © DEPARTAMENT ID 2008

Scorul la testul de calcul aritmetic			Scorul la testul de raționament verbal			Produsul abaterilor de la medie
X	$(x-m_x)$	$(x-m_x)^2$	**Y**	$(y-m_y)$	$(y-m_y)^2$	$(x-m_x)*(y-m_y)$
25	-4.63	21.44	28	-1.88	3.53	8.70
32	2.37	5.62	27	-2.88	8.29	-6.83
40	10.37	107.54	41	11.12	123.65	115.31
29	-0.63	0.40	34	4.12	16.97	-2.60
31	1.37	1.88	25	-4.88	23.81	-6.69
16	-13.63	185.78	19	-10.88	118.37	148.29
28	-1.63	2.66	26	-3.88	15.05	6.32
36	6.37	40.58	39	9.12	83.17	58.09
Σ= 237		Σ=365.88	Σ= 239		Σ=392.88	Σ=320,63
m_X= 29.63			m_Y= 29.88			
s_X = 7.23			s_Y = 7.49			

Pentru calcularea coeficientului de corelație am ales, de data aceasta, formula 3.30, prin care, înlocuind valorile, obținem valoarea coeficientului de corelație:

$$r = \frac{\sum(X-m_x)*(Y-m_y)}{N*s_x*s_y} = \frac{320.63}{8*7.23*7.49} = \frac{320.63}{433.22} = +0.74$$

Graficul *scatterplot* pentru datele din exemplu este corespunzător unei asocieri pozitive între cele două variabile, norul de puncte urmând o elipsă cu diagonala mare pe direcția stânga jos-dreapta sus:



Decizia statistică

Valoarea calculată a lui *r* (+0.74) este mai mare decât valoarea critică (+0.62), fapt care îndreptățește respingerea ipotezei de nul. Ca urmare, acceptăm ca semnificativ coeficientul de corelație obținut. Datele cercetării susțin ipoteza că între scorurile celor două teste există o legătură pozitivă semnificativă[20].

Corelație şi cauzalitate

Coeficientul de corelație ne oferă informații despre modul în care variază valorile a două variabile, una în raport cu cealaltă. Ca urmare, coeficientul de corelație nu are o semnificație cauzală decât dacă cele două variabile au fost măsurate într-un context care probează cauzalitatea. Iar acest lucru se petrece numai în condiții de experiment.

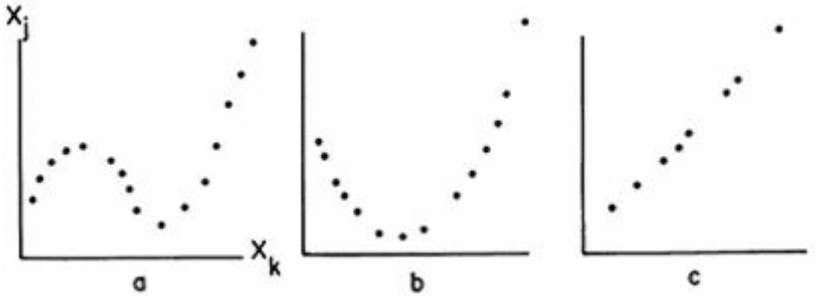
Natura liniară a corelației Pearson

Trebuie să reținem faptul că ceea ce exprimă *r* este intensitatea corelației liniare, adică măsura în care norul de puncte reprezentat de intersecția valorilor perechi ale celor două variabile poate fi reprezentat de o linie dreaptă. Asocierea de tip liniar este însă doar una dintre formele de aproximare a legăturii dintre variabile. În realitate, uneori, corelația dintre două variabile are o formă care se abate de la modelul rectiliniu

[20] În mod uzual, valorile lui ***r*** se raportează cu două zecimale, chiar dacă valorile tabelare şi cele calculate de programele statistice sunt cu mai mult de două zecimale.

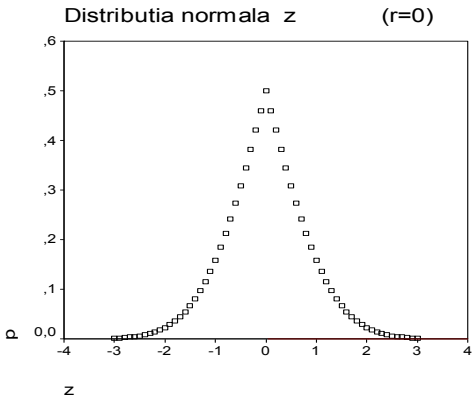
Copyright © DEPARTAMENT ID 2008

(are o formă curbă). Dacă privim imaginile de mai jos, putem observa câteva tipuri posibile de curbe de corelație. Figurile *a* şi *b* exprimă corelații perfecte, dar care se supun unui model curbiliniu, în timp ce figura *c* reprezintă o corelație perfectă rectilinie.



Relațiile curbilinii sunt calculate pe baza altor proceduri decât coeficientul Pearson (*r*), dar acestea nu fac de regulă obiectul de studiu al unei introduceri în statistica aplicată. Să reținem totuşi că, dacă am calcula un coeficient *r* pentru distribuțiile din figurile a şi b, atunci valoarea acestora ar fi foarte mică şi, cel mai probabil, nesemnificativă, în ciuda asocierii grafice evidente a valorilor lor.

Pentru a înțelege şi mai bine acest fapt, oferim un exemplu ilustrativ. Am introdus valorile lui ***z*** şi probabilitățile corespunzătoare lor de pe curba normală, într-un program de prelucrări statistice. Forma normală a curbei obținute ne indică faptul că, dinspre partea stângă a acesteia, valorile *z* devin din ce în ce mai mici (în valoare absolută), corespunzător cu creşterea probabilității, până la mijlocul curbei, unde *z*=0, iar probabilitatea este maximă. Mergând mai departe, spre dreapta, valorile lui *z* încep să crească, concomitent cu reducerea probabilității. Coeficientul de corelație calculat pentru un eşantion de date ale celor două variabile statistice este *r*=0, iar imaginea *scatterplot* a relației dintre ele este prezentată în figura alăturată[21]:



Aşa cum se observă, deşi *r*=0, ceea ce indică absența oricărei corelații liniare între variabile, curba de distribuție arată o corelație curbilinie perfectă.

Din fericire, astfel de situații sunt relativ rare în realitate, modelul corelației liniare fiind adecvat pentru un mare număr de relații dintre variabilele naturale, incluzându-le şi pe cele psihologice. Atunci când există suspiciuni consistente cu privire la natura liniară a legăturii dintre variabile, se pot efectua anumite transformări care să le aducă în cadrul unei variații liniare (de exemplu, extragerea radicalului sau logaritmarea variabilelor). Atunci când se raportează un coeficient de corelație fără a se preciza caracterul liniar sau curbiliniu, vom considera că acesta se referă la corelația liniară.

Exemplul dat ne sugerează faptul că graficul *scatterplot* oferă informații suplimentare semnificative şi, din acest motiv, este recomandabilă realizarea acestuia de fiecare dată când utilizăm testul de corelație Pearson. Un argument spectaculos în sprijinul acestui aspect ne este oferit de Anscombe (1973), care a realizat cele patru seturi de date din tabelul de mai jos:

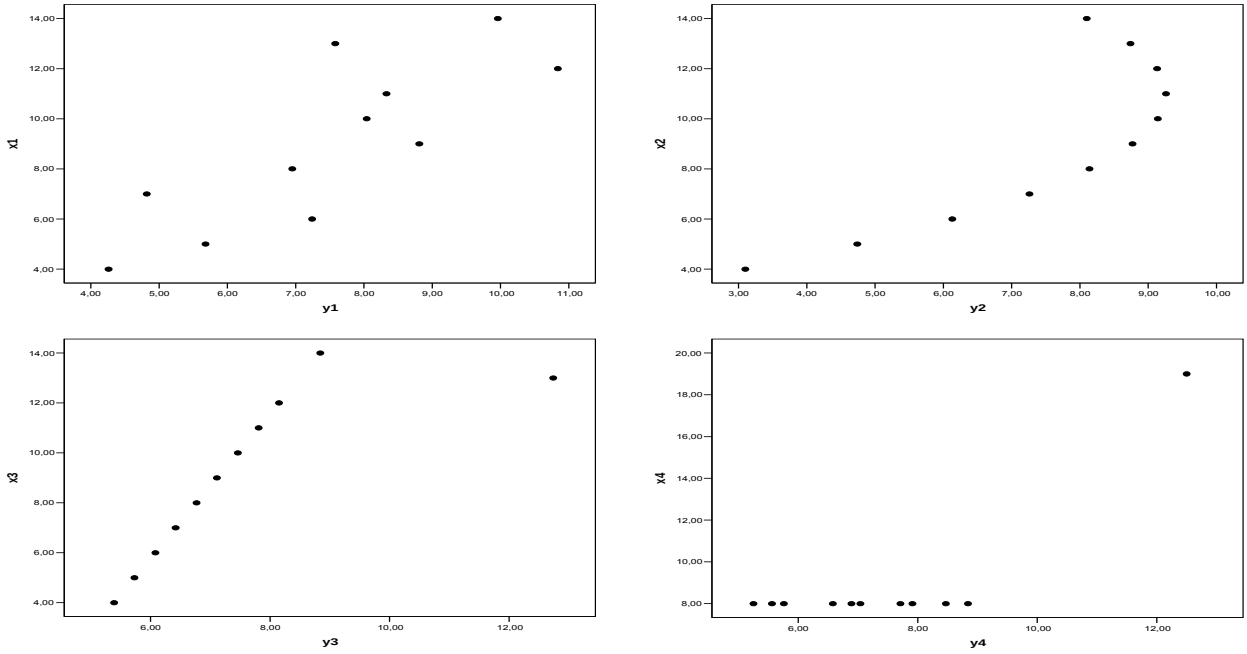
Setul #1		**Setul #2**		**Setul #3**		**Setul #4**	
X_1	Y_1	X_2	Y_2	X_3	Y_3	X_4	Y_4
10,00	8,04	10,00	9,14	10,00	7,46	8,00	6,58
8,00	6,95	8,00	8,14	8,00	6,77	8,00	5,76
13,00	7,58	13,00	8,74	13,00	12,74	8,00	7,71
9,00	8,81	9,00	8,77	9,00	7,11	8,00	8,84
11,00	8,33	11,00	9,26	11,00	7,81	8,00	8,47

[21] Exemplul se bazează pe un eşantion de 61 de perechi de valori, selectate de pe toată plaja distribuției *z*

Copyright © DEPARTAMENT ID 2008

14,00	9,96	14,00	8,10	14,00	8,84	8,00	7,04
6,00	7,24	6,00	6,13	6,00	6,08	8,00	5,25
4,00	4,26	4,00	3,10	4,00	5,39	19,00	12,50
12,00	10,84	12,00	9,13	12,00	8,15	8,00	5,56
7,00	4,82	7,00	7,26	7,00	6,42	8,00	7,91
5,00	5,68	5,00	4,74	5,00	5,73	8,00	6,89

Coeficienții de corelație dintre cele patru perechi de variabile (X_1-Y_1, X_2-Y_2, X_3-Y_3; X_4-Y_4) sunt identici: r=0.81. Şi totuşi, dacă sunt analizate reprezentările *scatterplot* pentru fiecare dintre cele patru perechi de variabile, imaginile ne oferă diferențe importante cu privire la natura reală a relației dintre ele:



În cazul perechii X_3-Y_3, o valoare extremă a redus coeficientul de corelație, iar în cazul perechii de variabile X_4-Y_4, unde corelația ar fi fost nulă, ea este generată de o singură valoare extremă. Desigur, astfel de efecte apar cu precădere în cazul eşantioanelor de volum mic, dar grija pentru valorile extreme trebuie menținută în toate cazurile.

Mărimea efectului coeficientului de crelație

Spre deosebire de testele *t*, introduse anterior, valoarea testului *r* este interpretabilă prin ea însăşi, exprimând intensitatea asocierii dintre variabile. Aşa cum am spus deja, avem o corelație perfectă atunci când *r* este egal cu +1 sau –1. Valoarea obținută pe exemplul nostru (+0.74) este destul de apropiată de +1. Desigur, +0.74 este mai puțin decât +1, dar şi mai mult decât, să zicem, +0.32. O asemenea interpretare, deşi absolut corectă, nu poate fi satisfăcătoare. Se simte necesitatea de a avea un criteriu de valorizare a cuantificării numerice a corelației. De-a lungul timpului au fost propuse diverse astfel de scale de valorizare, prin atribuirea unor calificative coeficienților de corelație, în funcție de mărimea lor. Această problemă comportă multe discuții, iar soluțiile oferite de diferiți autori sunt deseori diferite. Ca regulă generală, toți autorii sunt de acord că valorile mai mici de ±0.1 ale coeficienților de corelație trebuie să fie considerate „neglijabile", chiar şi atunci când ating pragul de semnificație statistică.

Oferim, cu caracter orientativ, modelul de descriere propus de Hopkins (2000) cu privire la interpretarea valorilor coeficienților de corelație:

Coeficientul de corelație	**Descriptor**
0.0-0.1	Foarte mic, neglijabil, nesubstanțial
0.1-0.3	Mic, minor
0.3-0.5	Moderat, mediu
0.5-0.7	Mare, ridicat, major

Copyright © DEPARTAMENT ID 2008

0.7-0.9	Foarte mare, foarte ridicat
0.9-1	Aproape perfect, descrie relația dintre două variabile practic indistincte

O altă variantă de interpretare a mărimii efectului recunoscută este cea propusă de Davis (citat de Kotrlik şi Williams, 2003):

0.70 →	asociere foarte puternică
0.50 – 0.69	asociere substanțială
0.30 – 0.49	asociere moderată
0.10 – 0.29	asociere scăzută
0.01 – 0.09	asociere neglijabilă

Înaintea oricărui calificativ însă, prima condiție pentru a lua în considerare existența unei corelații între două variabile rămâne atingerea pragului de semnificație (*alfa*). Dacă valoarea lui *r* este mai mică decât *r* critic (corespunde unui nivel *p* mai mare de 0.05, sau decât alt prag legitim decis de cercetător), existența unei corelații nu poate fi luată în considerare, indiferent de mărimea coeficientului *r* Pearson. Aceasta, deoarece nu avem temei pentru a accepta că se îndepărtează suficient de o valoare care ar fi putut decurge prin jocul hazardului.

În cele din urmă ce trebuie să luăm în considerare, semnificația sau intensitatea asocierii? Desigur, răspunsul este unul relativ. Dacă finalitatea studiului este aceea de a lua decizii, ca în cazul selecției de personal, de exemplu, se vor căuta valori cât mai mari ale coeficientului de corelație (*r*). Dar dacă obiectivul este preponderent teoretic, de a pune în evidență relații „ascunse" între variabile, atunci, indiferent de mărimea lor, coeficienții de corelație vor fi luați în considerare (desigur, dacă sunt mai mari de 0.1).

Coeficientul de determinare

Valorile lui ***r*** trebuie considerate pe o scală ordinală. Ca urmare, nu este corect să afirmăm că un coeficient de corelație de 0.40 este de două ori mai mare decât un altul de 0.20. Dacă dorim să comparăm în mod direct doi coeficienți de corelație trebuie să ridicăm valorile lui *r* la pătrat (r^2), obținând astfel ceea ce se numeşte *coeficient de determinare* (prezentat în programele statistice şi ca „*r squared"*). Acesta este considerat un indicator mai adecvat al mărimii efectului, deoarece ia valori sensibili mai mici decât cele ale coeficientului de corelație. Pentru exemplul nostru, coeficientul de determinare este 0.74^2=0.55. Transformat în procente, acest rezultat se interpretează astfel: „*55% din variația (împrăştierea) uneia dintre cele două variabile este determinată de variația celeilalte variabile*". Sau, altfel spus, cele două variabile au in comun 55% din variația care le caracterizează, ceea ce înseamnă că 45% din variabilitatea lor provine din alte surse. Atenție, interpretarea procentuală, în maniera prezentată, este valabilă numai pentru coeficientul de determinare. Coeficientul de corelație (*r*) nu poate fi interpretat în nici un caz sub formă procentuală!

Cohen (citat de Kotrlik şi Williams, 2003) a propus următoarea regulă de evaluare a mărimii coeficentului de determinare ca indice de mărime a efectului în cazul corelației:

r^2 (Cohen)	0.0196	efect mic
	0.1300	efect mediu
	0.2600	efect mare

Vom observa că valorile lui r corespunzătoare celor trei praguri ale lui r2 sunt 0.14, 0.36 şi, respectiv, 0.50, ceea ce este în concordanță cu recomandările de mai sus pentru interpretarea lui r.

Să reținem că mărimea efectului, care este, de fapt, însăşi mărimea coeficientului r, depinde de două elemente principale:

- Caracterul liniar al relației dintre variabile. O componentă curbilinie a asocierii va conduce la valori mai mici ale coeficientului de corelație. Graficul *scatterplot* ne poate ajuta la evidențierea acestui aspect.
- Variabilitatea distribuțiilor comparate. Dacă variabilele cercetate au o împrăştiere redusă, acest fapt limitează posibilitatea de a obține valori ridicate pentru *r*.

Mărimea eşantionului are efect doar asupra puterii testului (eşantioanele mari conduc mai uşor la atingerea pragului de semnificație), dar nu au un efect important asupra mărimii lui *r*.

Copyright © DEPARTAMENT ID 2008

Limitele de încredere pentru coeficientul de corelație *r*

Atunci când calculăm coeficientul de corelație pentru valorile măsurate pe un eşantion, o facem, desigur, cu scopul de a avea o estimare asupra gradului în care cele două variabile au o variație comună la nivelul întregii populații. Deoarece calcularea corelației pe „valorile populației" este practic imposibilă, tot ce putem face este să o estimăm, cu o anumită marjă de eroare, prin utilizarea corelației pe eşantion. Astfel, în termeni formali, *r* (calculat pentru eşantion) este o estimare pentru ρ (*ro*), corelația „adevărată" la nivelul populației.

Calcularea limitelor de încredere

Construirea intervalelor de încredere pentru coeficientul de corelație la nivelul populației (ρ) nu este la fel de simplă ca în cazul altor valori statistice, dar se bazează pe acelaşi raționament fundamental: limitele de încredere se află în jurul unui punt de estimare (*r*) la care se adaugă sau se scade valoarea *r* critic înmulțită cu eroarea standard a estimării. Problemele specifice decurg din natura distribuției lui *r*. Atunci când valoarea corelației la nivelul populației este ρ=0, distribuția de eşantionare r_s (valorile lui *r* care ar fi calculate pe eşantioanele extrase din aceeaşi populație) formează o distribuție normală în jurul lui zero (dacă volumul eşantionului este suficient de mare). Dar dacă ρ=+0.7, distribuția lui r_s are o împrăştiere asimetrică în jurul lui acestei valori. Motivul este simplu: este mai mult „loc" pentru valori sub +0.7 decât peste această valoare, deoarece ştim că *r* ia valori între -1 şi +1. Cu cât estimarea pentru ρ este mai aproape de limitele teoretice ale lui *r*, cu atât distribuția r_s este mai asimetrică spre partea opusă. Această particularitate creează o piedică în transformarea coeficienților r_s în scoruri Z (cu majusculă, pentru a se evita confuzia cu scorurile *z* clasice), necesare construirii limitelor intervalului de încredere pentru ρ. Problema a fost rezolvată de Fisher, care a elaborat un algoritm pe baza căruia valorile r_s sunt transformate în valori Z, a căror arie de distribuție sub curba normală este cunoscută:

$$Z=0.5\ln[(1 + r)/(1 - r)] \qquad \text{(formula 3.31)}$$

Pentru a se evita aplicarea acestei formule relativ greoaie, se poate utiliza un tabel (vezi în anexă tabelul Fisher de transformare în Z a valorilor lui *r*) care, chiar dacă nu conține toate valorile intermediare, este suficient pentru a acoperi nevoile practice.

Să luăm ca exemplu valoarea coeficientului de corelație parțială obținut de noi mai sus: *r*=+0.74. Ne propunem să aflăm care sunt limitele de încredere ale acestei valori, adică să definim intervalul în care se poate afla valoarea reală a corelației la nivelul populației, cu o probabilitate asumată. De regulă, aşa cum ştim, această probabilitate asumată este de 0.05 sau, exprimată altfel, un nivel de încredere de 95%.

Practic, aflarea limitelor se face în felul urmîtor:

- Se transformă *r* calculat în valoare Z, citind tabela Fisher: în cazul nostru, pentru *r*=0.85 avem o valoare Z=0.9505 (dacă valoarea lui r nu se regăseşte ca atare în tabel, se poate face o medie a valorilor apropiate). Pe o distribuție normală, cum este distribuția de eşantionare Z, ştim că aproximativ 95% dintre valori se întind între -1.96 şi +1.96. Adică, pe o distanță de aproximativ două abateri standard în jurul mediei (abaterea standard a valorilor Z fiind 1).
- Se calculează eroarea standard a transformării Z, cu formula:

$$r_e = \frac{1}{\sqrt{N-3}} = \frac{1}{\sqrt{8-3}} = 0{,}447$$ unde N este volumul eşantionului

- Se calculează limitele superioară şi inferioară a intervalului: $\rho = r \pm z_{critic} * r_e$, adică:
  Limita superioară (Z): 0.9505+1.96*0.447=+1.826
  Limita inferioară (Z): 0.9505-1.96*0.447=+0.074

Limitele astfel calculate sunt exprimate în valori transformate Z, ori noi avem nevoie să ştim limitele în valori ale lui r. Pentru aceasta, facem acum transformarea inversă, citind valorile lui Z în tabela Fisher, corespunzătoare celor două limite de mai sus:

Limita superioara de încredere pentru *r*=+0.95
Limita inferioară de încredere pentru *r*=+0.07

În concluzie, valoarea adevărată (la nivelul populației) a corelației dintre cele două variabile, se află, cu o probabilitate de 95%, în intervalul cuprins între +0.07 şi +0.95. Limita inferioară este în apropierea unei

Copyright © DEPARTAMENT ID 2008

corelații egale cu 0, iar limita superioară în vecinătatea corelației perfecte, ceea ce ne arată o precizie de estimare scăzută. Acest fapt este normal, dacă avem în vedere mărimea redusă a eşantionului, care determină un nivel ridicat al erorii standard pentru r (prin faptul că se află la numitorul formulei).

Utilizarea limitelor de încredere

Dacă analizăm limitele intervalului de încredere astfel obținute, pentru exemplul nostru, trebuie să constatăm că ele sunt foarte mari, în zona valorilor pozitive, dar având limita inferioară destul de aproape de valoarea zero. Acest fapt conduce la concluzia că, deşi este atât mare şi semnificativ statistic, coeficientul obținut are o valoare mică de generalizare. Situația este generată, în acest caz, de volumul extrem de mic al eşantionului. Amplitudinea intervalului de încredere este direct dependentă de volumul eşantionului. Cu cât N este mai mare, cu atât valoarea erorii standard tinde să scadă, ceea ce aduce limitele intervalului de încredere mai aproape de valoarea calculată a lui *r*.

Să ne imaginăm că am efectuat un calcul de corelație pe 30 de subiecți şi am obținut r=0.30 (când semnul corelației nu este specificat, se consideră pozitiv). Limitele de încredere pentru acesta sunt între -0.07 şi +0.60, ceea ce arată că este nesemnificativ, dat fiind faptul că între cele două limite este şi valoarea zero, aceea care este vizată de ipoteza de nul. Faptul că limita inferioară este foarte aproape de valoarea zero (la numai 7 sutimi de ea), ne îndreptățeşte să credem că, prin mărirea volumului eşantionului de cercetare ar putea fi atins nivelul de semnificație statistic. Aceasta, deoarece în formula erorii standard a lui *r* volumul eşantionului se află la numitor şi, cu cât N va fi mai mare, cu atât valoarea lui r_e va fi mai mică, iar limitele intervalului de încredere pentru *r*, mai aproape de *r*.

Tabelul următor arată care sunt limitele pentru exemplul dat, dacă N ar creşte, progresiv, până la 100:

N	Pearson ***r***	Niv. de încredere (%)	Limite de încredere	
			inferioară	superioară
30	0,30	95	-0,07	0,60
40	0,30	95	-0,01	0,56
50	0,30	95	0,02	0,53
60	0,30	95	0,05	0,51
70	0,30	95	0,07	0,50
80	0,30	95	0,09	0,49
90	0,30	95	0,10	0,48
100	0,30	95	0,11	0,47

Utilitatea practică a acestor estimări de limite este dată de faptul că ne arată cu cât ar trebui să creştem volumul eşantionului pentru a obține un rezultat semnificativ al coeficientului de corelație dintre cele două variabile. Aşa cum se vede, dacă am creşte volumul eşantionului la 50 de subiecți, limita inferioară ar trece deja peste valoarea zero. Celelalte linii din tabel prezintă efectul de mărime al eşantionului în cazul creşterii lui N până la 100 de subiecți.

Semnificația diferenței dintre doi coeficienți de corelație

Să presupunem că într-o cercetare este evaluată corelația dintre extraversie şi agresivitate separat, pentru bărbați şi pentru femei, obținându-se o valoare *r*=0.50 pentru bărbați şi o valoare *r*=0.30 pentru femei, ambii coeficienți fiind semnficativi. În acest caz ne-am putea pune problema dacă cei doi coeficienți diferă semnificativ între ei, ceea ce ar însemna că relația dintre extraversie şi agresivitate este mai ridicată la bărbați decât la femei.

Diferența dintre doi coeficienți de corelație poate fi evaluată cu un test specific, care ia în considerare nu doar diferența dntre valorile *r*, ci şi mărimea eşantioanelor şi mărimea în sine a celor doi coeficienți. De exemplu, având în vedere că semnificația coeficienților de corelație depinde şi de mărimea eşantionului, înseamnă că o diferență de 0.1 între doi indici de corelație poate fi nesemnficativă dacă cei doi *r* sunt 0.15 şi 0.25, dar poate fi semnificativă dacă valorile *r* comparate sunt 0.80 şi 0.90.

Copyright © DEPARTAMENT ID 2008

Modul de calcul al semnificației dintre doi coeficienți de corelație va fi prezentat mai tîrziu, în secțiunea rezervată procedurilor SPSS.

Condiții pentru calcularea coeficientului de corelație Pearson

Pentru a putea utiliza în mod legitim calculul de corelație eşantionul trebuie să fie aleatoriu, iar cele două variabile (ambele măsurate pe scale de interval/raport) trebuie să aibă o distribuție care să nu se abată grav de la distribuția normală. Această condiție este cu atât mai importantă cu cât eşantionul este mai mic. O atenție aparte trebuie acordată valorilor excesive, prezența acestora putând avea efecte neaşteptate asupra valorii coeficientului de corelație (vezi exemplele lui Anscombe).

Utilizarea coeficientul de corelație

Analiza de corelație este una dintre cele mai uzuale proceduri statistice în cercetarea psihologică. Printre utilizările cele mai comune menționăm analiza *consistenței* şi *validității* testelor psihologice. *Consistența* se referă la gradul în care un instrument de evaluare se concentrează asupra unei anumite realități psihice. *Validitatea*, se referă la faptul dacă ceea ce presupune că măsoară un instrument psihologic este măsurat cu adevărat (de exemplu, o scală de anxietate măsoară cu adevărat anxietatea?).

Din cele prezentate, rezultă că putem utiliza coeficientul atunci când avem serii perechi de distribuții. Pentru o mai bună înțelegere, se cuvine să facem câteva aprecieri comparative cu testul t pentru eşantioane dependente. Testul ***t*** pentru eşantioane dependente, se aplică atunci când măsurăm o anumită variabilă în două situații diferite (de ex. *înainte/după*), ceea ce presupune aceeaşi unitate de măsură. Coeficientul de corelație poate fi aplicat atât pentru variabile măsurate cu aceeaşi unitate de măsură cât şi pentru variabile exprimate în unități de măsură diferite. Aceasta deoarece formula de calcul ia în considerare expresia standardizată a valorilor (corurile ***z***). Întrebarea este, când utilizăm unul sau altul dintre cele două teste? Răspunsul ține de scopul pe care ni-l propunem. Dacă dorim să punem în evidență diferența dintre valorile medii ale variabilelor, vom aplica testul ***t*** pentru eşantioane dependente. Dacă ne interesează intensitatea variației concomitente a variabilelor, vom utiliza coeficientul de corelație.

Coeficientul de corelație Pearson nu este singurul test al asocierii variabilelor. Există o varietate de teste de corelație, utilizate pentru situațiile în care variabilele cercetate sunt măsurate, fiecare, pe oricare dintre scalele de măurare.

Publicarea rezultatului corelației

Raportarea coeficienților de corelație va cuprinde, pe lângă indicatorii statistici descriptivi ai variabiulelor (medii, abateri standard, indicatorii simetriei şi aplatizării), volumul eşantionului, valoarea lui *r*, nivelul de semnificație şi coeficientul de determinare (r^2). Prezentarea limitelor de încredere nu este uzuală, poate şi pentru că programele statistice obişnuite nu le oferă, dar calcularea şi includerea lor în documentul cercetării este de dorit.

Pentru exemplul de mai sus, o prezentare narativă a rezultatului ar putea arăta astfel:
„*A fost evaluată performanța la un test de calcul aritmetic şi la unul de raționament verbal logic, pentru un eşantion de 6 subiecți. Scorurile mari se referă la performanțe ridicate. Media scorului la primul test a fost de m=29.63 (s=6.76), iar la al doilea m=29.88 (s=7.01). Am obținut o corelație semnificativă între cele două performanțe, r=0.74 (r^2=0.55), p<0.05, bilateral. Limitele de încredere pentru coeficientul r (95%) sunt cuprinse între +0.07 şi +0.95.*"

NOTĂ: Se precizează neapărat semnificația valorilor variabilelor în raport de mărimea lor, pentru a se putea aprecia corect natura relației dintre variabile.

Sarcina de lucru 2.7

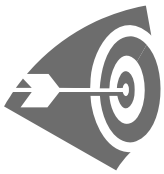
Un eşantion de 10 subiecți a fost testat cu un test de inteligență de două ori, pentru a se verifica stabilitatea test-retest a instrumentului. Rezultatele, marcate cu I şi II, sunt prezentate în tabelul următor:

Copyright © DEPARTAMENT ID 2008

I	II
110	105
100	108
120	110
90	95
108	105
115	125
122	118
110	116
127	118
118	126

1. Enunțați ipoteza statistică
2. Enunțați ipoteza cercetării
3. Definiți populațiile cercetării
4. Definiți criteriile de decizie statistică
5. Calculați coeficientul de corelație *Pearson*
6. Enunțați decizia statistică
7. Enunțați concluzia cercetării.

Verificați răspunsurile corecte.



Rezumatul unității de învățare

- Testele statistice verifică adevărul ipotezelor de cercetare prin testarea adevărului ipotezei statistice (de nul). Atunci când ipoteza de nul este respinsă, ipoteza cercetării este confirmată. Atunci când ipoteza de nul este acceptată, ipoteza cercetării nu se confirmă.
- Testul z pentru un singur eşantion compară scorul z al unui eşantion cu media populației din care provine, pentru a testa dacă eşantionul face pare din acea populație (numită „de nul") sau nu.
- Condiția limitativă a utilizării testului z pentru un eşantion este cunoaşterea mediei populației.
- Testul t pentru un eşantion este similar cu testul z, dar se aplică în cazul eşantioanelor mai mici de 30 de subiecți.
- Testul t pentru eşantioane independente verifică semnificația diferenței dintre mediile a două eşantioane formate din subiecți diferiți, care au fost evaluate în condiții distincte.
- Respingerea ipotezei de nul în acest caz, înseamnă acceptarea ipotezei că cele două medii sunt diferite, ceea ce este echivalent cu acceptarea ipotezei cercetării care afirmă existența unei relații între condiția testării (variabila independentă, măsurată pe scală nominală, dihotomică) şi rezultatul măsurat prin variabila dependentă (măsurată pe scală cantitativă).
- Analiza de varianță (ANOVA) testează diferența dintre mediile a mai mult de două medii obținute pe eşantioane independente.
- Testul ANOVA se bazează pe evidențierea diferenței dintre medii prin analiza variabilității lor.
- ANOVA este necesară în cazul comparării a mai mult de două medii deoarece compararea acestora două câte două este nepermisă, ca urmare a acumulării excesive de eroare de tip I.
- Testul t pentru diferența mediilor a două eşantioane dependente vizează situațiile în care aceiaşi subiecți au fost evaluați cu acelaşi instrument în situații diferite. Variabila independentă este reprezentată de condiția în care are loc măsurarea, iar variabila dependentă este caracteristica care face obiectul măsurării, fiind exprimată pe scală cantitativă.
- Coeficientul de corelație Pearson testează intensitatea asocierii de tip liniar dintre două variabile măsurate pe aceiaşi subiecți, în condiții diferite sau cu instrumente diferite. Coeficientul de corelație

Copyright © DEPARTAMENT ID 2008

variază între -1 şi +1, care exprimă corelații maxime (negativă, respectiv pozitivă). Un coeficient de corelație egal cu 0 semnifică absența oricărei asocieri între cele două variabile.

Răspunsuri la sarcinile de lucru
Sarcina de lucru 2.3

z critic pentru alfa=0.05 unilateral este 1.65
$z_{calculat}$ =-0,16, mai mare decât z_{critic}

Decizia statistică: Se acceptă ipoteza de nul.
Concluzia cercetării: Ipoteza cercetării nu se confirmă. Medicamentul respectiv nu produce bradicardie (reducerea pulsului).

Sarcina de lucru 2.4

1. Terapia împotriva fobiei are un efect?
2. Terapia antifobică reduce simptomele fobice.
3. Terapia antifobică nu reduce simptomele fobice.
4. t critic=-2.04
5. Aplicând formula testului t pentru dispersii diferite obținem:

$$t = \frac{27.2 - 34.4}{\sqrt{\frac{s_1^{\,2}}{N1} + \frac{s_2^{\,2}}{N2}}} = \frac{-7.2}{3.75} = -1.92$$

6. Se acceptă ipoteza de nul deoarece t calculat (-1.92)este mai mic decât t critic (-2.04)
7. Ipoteza cercetării nu se confirmă. Rezultatele nu susțin ipoteza că tratamentul împotriva fobiei are vreun efect.

Sarcina de lucru 2.5

1. F critic pentru 2 cu 15 grade de libertate =3.6823
2. F calculat= 16,4
3. Se respinge ipoteza de nul
4. 18 subiecți au fost repartizați aleatoriu în 3 grupuri, fiecare primind bomboane de culori diferite (roşu, verde, galben). Mediile de consum pentru fiecare grup în parte au fost: Cele trei grupuri au consumat în medie, m1=1.5; m2=4.67 şi m3=1.5. Analiza de varianță unifactorială a relevat o diferență semnificativă între medii F(2,15)=16.4; *p≤0.05. Rezultatele indică faptul că cea mai preferată culoare de bomboane este roşu.*

Sarcina de lucru 2.6

1. Ipoteza cercetării: Stresul temporal diminuează performanța de calcul numeric. Ipoteza de nul: Stresul temporal nu are nici un efect asupra performanței de calcul.
2. t critic=2.36
3. t calculat=2.29
4. Se acceptă ipoteza de nul
5. Se infirmă ipoteza cercetării
6. Un eşantion de 8 subiecți a fost supus unui test de calcul aritmetic în condiții de stres temporal şi în condiții fără stres. Nivelul performanței în condiții de stres a scăzut (m=76.88) față de condiții lipsite de stres (m=85.38). Diferența nu s-a dovedit totuşi semnificativă statistic t(7)=2.29, p>0.05, pentru alfa=0.05, bilateral.

Copyright © DEPARTAMENT ID 2008

Observație: *În acest caz, rezultatul nesemnificativ al testului t este determinat de numărul mic de subiecți. Așa cum se observă, valoarea calculată a lui t este foarte aproape de t critic.*

Sarcina de lucru 2.7

1. Ipoteza cercetării: Scorurile obținute de subiecți la cele două aplicări vor fi concordante.
2. Ipoteza statistică: Scorurile nu vor fi concordante
3. Populațiile cercetării sunt populația subiecți la prima aplicare şi populația de subiecți la a doua aplicare a testului.
4. Criteriile de decizie statistică: alfa 0.05 bilateral; df=10-2=8, r critic=063
5. r calculat=0.73
6. Decizia statistică: Se respinge ipoteza de nul.
7. Concluzia cercetării: Scorurile sunt concordante de la o aplicare la alta.

Lucrarea de evaluare nr. 2.2

Lucrarea de evaluare va fi publicată pe portal. Termenul limită de trimitere este preziua tutorialului. După această dată lucrările trimise nu vor mai fi luate în considerare.

Bibliografie minimală

Bibliografia de bază

- Marian Popa, (2008), Statistică pentru psiholgie. Teorie □i aplica□ii SPSS, Polirom
- Pagina web a cursului, la adresa: ***www.mpopa.ro***

Bibliografie suplimentară

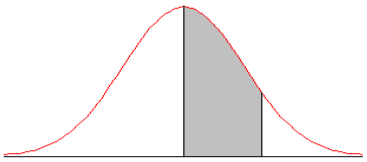
- Clocotici, V., & Stan, A. (2000). Statistica aplicata in psihologie. Iasi: Polirom, (selectiv)
- Rotaru, T. (coord.). (1999). Metode statistice aplicate in stiintele sociale. Iasi: Polirom. (selectiv)
- Vasilescu, I. P. (1992). Statistica informatizata pentru stiinte despre om (Vol. 1-2). Bucuresti: Editura militara., (selectiv)

Copyright © DEPARTAMENT ID 2008

ANEXE – TABELE STATISTICE

Anexa 1. Tabelul distribuției valorilor sub curba normală z

Sursa: http://brd1.ucsc.edu/exp_design/Z_table.htm



0 z

Valorile din tabel indică probabilitatea dintre 0 şi z.

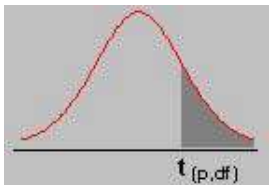
Z	0	0.01	0.02	0.03	0.04	0.05	0.06	0.07	0.08	0.09
0	0.00000	0.00399	0.00798	0.01197	0.01595	0.01994	0.02392	0.02790	0.03188	0.03586
0.1	0.03983	0.04380	0.04776	0.05172	0.05567	0.05962	0.06356	0.06749	0.07142	0.07535
0.2	0.07926	0.08317	0.08706	0.09095	0.09483	0.09871	0.10257	0.10642	0.11026	0.11409
0.3	0.11791	0.12172	0.12552	0.12930	0.13307	0.13683	0.14058	0.14431	0.14803	0.15173
0.4	0.15542	0.15910	0.16276	0.16640	0.17003	0.17364	0.17724	0.18082	0.18439	0.18793
0.5	0.19146	0.19497	0.19847	0.20194	0.20540	0.20884	0.21226	0.21566	0.21904	0.22240
0.6	0.22575	0.22907	0.23237	0.23565	0.23891	0.24215	0.24537	0.24857	0.25175	0.25490
0.7	0.25804	0.26115	0.26424	0.26730	0.27035	0.27337	0.27637	0.27935	0.28230	0.28524
0.8	0.28814	0.29103	0.29389	0.29673	0.29955	0.30234	0.30511	0.30785	0.31057	0.31327
0.9	0.31594	0.31859	0.32121	0.32381	0.32639	0.32894	0.33147	0.33398	0.33646	0.33891
1	0.34134	0.34375	0.34614	0.34849	0.35083	0.35314	0.35543	0.35769	0.35993	0.36214
1.1	0.36433	0.36650	0.36864	0.37076	0.37286	0.37493	0.37698	0.37900	0.38100	0.38298
1.2	0.38493	0.38686	0.38877	0.39065	0.39251	0.39435	0.39617	0.39796	0.39973	0.40147
1.3	0.40320	0.40490	0.40658	0.40824	0.40988	0.41149	0.41309	0.41466	0.41621	0.41774
1.4	0.41924	0.42073	0.42220	0.42364	0.42507	0.42647	0.42785	0.42922	0.43056	0.43189
1.5	0.43319	0.43448	0.43574	0.43699	0.43822	0.43943	0.44062	0.44179	0.44295	0.44408
1.6	0.44520	0.44630	0.44738	0.44845	0.44950	0.45053	0.45154	0.45254	0.45352	0.45449
1.7	0.45543	0.45637	0.45728	0.45818	0.45907	0.45994	0.46080	0.46164	0.46246	0.46327
1.8	0.46407	0.46485	0.46562	0.46638	0.46712	0.46784	0.46856	0.46926	0.46995	0.47062
1.9	0.47128	0.47193	0.47257	0.47320	0.47381	0.47441	0.47500	0.47558	0.47615	0.47670
2	0.47725	0.47778	0.47831	0.47882	0.47932	0.47982	0.48030	0.48077	0.48124	0.48169
2.1	0.48214	0.48257	0.48300	0.48341	0.48382	0.48422	0.48461	0.48500	0.48537	0.48574
2.2	0.48610	0.48645	0.48679	0.48713	0.48745	0.48778	0.48809	0.48840	0.48870	0.48899
2.3	0.48928	0.48956	0.48983	0.49010	0.49036	0.49061	0.49086	0.49111	0.49134	0.49158
2.4	0.49180	0.49202	0.49224	0.49245	0.49266	0.49286	0.49305	0.49324	0.49343	0.49361
2.5	0.49379	0.49396	0.49413	0.49430	0.49446	0.49461	0.49477	0.49492	0.49506	0.49520
2.6	0.49534	0.49547	0.49560	0.49573	0.49585	0.49598	0.49609	0.49621	0.49632	0.49643
2.7	0.49653	0.49664	0.49674	0.49683	0.49693	0.49702	0.49711	0.49720	0.49728	0.49736
2.8	0.49744	0.49752	0.49760	0.49767	0.49774	0.49781	0.49788	0.49795	0.49801	0.49807
2.9	0.49813	0.49819	0.49825	0.49831	0.49836	0.49841	0.49846	0.49851	0.49856	0.49861
3	0.49865	0.49869	0.49874	0.49878	0.49882	0.49886	0.49889	0.49893	0.49896	0.49900
3.1	0.49903	0.49906	0.49910	0.49913	0.49916	0.49918	0.49921	0.49924	0.49926	0.49929
3.2	0.49931	0.49934	0.49936	0.49938	0.49940	0.49942	0.49944	0.49946	0.49948	0.49950
3.3	0.49952	0.49953	0.49955	0.49957	0.49958	0.49960	0.49961	0.49962	0.49964	0.49965
3.4	0.49966	0.49968	0.49969	0.49970	0.49971	0.49972	0.49973	0.49974	0.49975	0.49976
3.5	0.49977	0.49978	0.49978	0.49979	0.49980	0.49981	0.49981	0.49982	0.49983	0.49983
3.6	0.49984	0.49985	0.49985	0.49986	0.49986	0.49987	0.49987	0.49988	0.49988	0.49989
3.7	0.49989	0.49990	0.49990	0.49990	0.49991	0.49991	0.49992	0.49992	0.49992	0.49992
3.8	0.49993	0.49993	0.49993	0.49994	0.49994	0.49994	0.49994	0.49995	0.49995	0.49995
3.9	0.49995	0.49995	0.49996	0.49996	0.49996	0.49996	0.49996	0.49996	0.49997	0.49997
4	0.49997	0.49997	0.49997	0.49997	0.49997	0.49997	0.49998	0.49998	0.49998	0.49998
4.1	0.49998	0.49998	0.49998	0.49998	0.49998	0.49998	0.49998	0.49998	0.49999	0.49999
4.2	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999

Copyright © DEPARTAMENT ID 2008

4.3	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999	0.49999
4.4	0.49999	0.49999	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
4.5	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
4.6	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
4.7	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
4.8	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
4.9	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
5	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
5.1	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
5.2	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
5.3	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
5.4	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
5.5	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
5.6	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
5.7	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
5.8	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
5.9	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000
6	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000	0.50000

Copyright © DEPARTAMENT ID 2008

Anexa 2. Tabelul valorilor critice pentru distribuția *t* Student (unilateral)
(sursa: http://www.psychology.ilstu.edu/psy138/tables.html)



df	alfa=0.10	alfa=0.05	alfa=0.025	alfa=0.01	alfa=0.005	alfa=0.0005
1	3.078	6.314	12.706	31.821	63.657	636.620
2	1.886	2.920	4.303	6.965	9.925	31.598
3	1.638	2.353	3.182	4.541	5.841	12.924
4	1.533	2.132	2.776	3.747	4.604	8.610
5	1.476	2.015	2.571	3.365	4.032	6.869
6	1.440	1.943	2.447	3.143	3.707	5.959
7	1.415	1.895	2.365	2.998	3.499	5.408
8	1.397	1.860	2.306	2.896	3.355	5.041
9	1.383	1.833	2.262	2.821	3.250	4.781
10	1.372	1.812	2.228	2.764	3.169	4.587
11	1.363	1.796	2.201	2.718	3.106	4.437
12	1.356	1.782	2.179	2.681	3.055	4.318
13	1.350	1.771	2.160	2.650	3.102	4.221
14	1.345	1.760	2.145	2.624	2.977	4.140
15	1.341	1.753	2.131	2.602	2.947	4.073
16	1.337	1.746	2.120	2.583	2.921	4.015
17	1.333	1.740	2.110	2.567	2.898	3.965
18	1.330	1.734	2.101	2.552	2.878	3.922
19	1.328	1.729	2.093	2.539	2.861	3.883
20	1.325	1.725	2.086	2.528	2.845	3.850
21	1.323	1.721	2.080	2.528	2.831	3.819
22	1.321	1.717	2.074	2.508	2.819	3.792
23	1.319	1.714	2.069	2.500	2.807	3.767
24	1.318	1.711	2.064	2.492	2.797	3.745
25	1.316	1.708	2.060	2.485	2.787	3.725
26	1.315	1.706	2.056	2.479	2.779	3.707
27	1.314	1.703	2.052	2.473	2.771	3.690
28	1.313	1.701	2.048	2.467	2.763	3.674
29	1.311	1.699	2.045	2.462	2.756	3.659
30	1.310	1.697	2.042	2.457	2.750	3.646
40	1.303	1.684	2.021	2.423	2.704	3.551
60	1.296	1.671	2.000	2.390	2.660	3.460
120	1.289	1.658	1.980	2.358	2.617	3.373
∞	1.282	1.645	1.960	2.326	2.576	3.291

Copyright © DEPARTAMENT ID 2008

Anexa 3. Tabelul parțial al distribuției F pentru α=0.05 [22]

(Sursa: Electronic Textbook, STATSOFT, Copyright StatSoft, Inc., 1984-1999)

df intragrup (within)	df intergrup (between)									
	1	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
1	161.4476	199.5000	215.7073	224.5832	230.1619	233.9860	236.7684	238.8827	240.5433	241.8817
2	18.5128	19.0000	19.1643	19.2468	19.2964	19.3295	19.3532	19.3710	19.3848	19.3959
3	10.1280	9.5521	9.2766	9.1172	9.0135	8.9406	8.8867	8.8452	8.8123	8.7855
4	7.7086	6.9443	6.5914	6.3882	6.2561	6.1631	6.0942	6.0410	5.9988	5.9644
5	6.6079	5.7861	5.4095	5.1922	5.0503	4.9503	4.8759	4.8183	4.7725	4.7351
6	5.9874	5.1433	4.7571	4.5337	4.3874	4.2839	4.2067	4.1468	4.0990	4.0600
7	5.5914	4.7374	4.3468	4.1203	3.9715	3.8660	3.7870	3.7257	3.6767	3.6365
8	5.3177	4.4590	4.0662	3.8379	3.6875	3.5806	3.5005	3.4381	3.3881	3.3472
9	5.1174	4.2565	3.8625	3.6331	3.4817	3.3738	3.2927	3.2296	3.1789	3.1373
10	4.9646	4.1028	3.7083	3.4780	3.3258	3.2172	3.1355	3.0717	3.0204	2.9782
11	4.8443	3.9823	3.5874	3.3567	3.2039	3.0946	3.0123	2.9480	2.8962	2.8536
12	4.7472	3.8853	3.4903	3.2592	3.1059	2.9961	2.9134	2.8486	2.7964	2.7534
13	4.6672	3.8056	3.4105	3.1791	3.0254	2.9153	2.8321	2.7669	2.7144	2.6710
14	4.6001	3.7389	3.3439	3.1122	2.9582	2.8477	2.7642	2.6987	2.6458	2.6022
15	4.5431	3.6823	3.2874	3.0556	2.9013	2.7905	2.7066	2.6408	2.5876	2.5437
16	4.4940	3.6337	3.2389	3.0069	2.8524	2.7413	2.6572	2.5911	2.5377	2.4935
17	4.4513	3.5915	3.1968	2.9647	2.8100	2.6987	2.6143	2.5480	2.4943	2.4499
18	4.4139	3.5546	3.1599	2.9277	2.7729	2.6613	2.5767	2.5102	2.4563	2.4117
19	4.3807	3.5219	3.1274	2.8951	2.7401	2.6283	2.5435	2.4768	2.4227	2.3779
20	4.3512	3.4928	3.0984	2.8661	2.7109	2.5990	2.5140	2.4471	2.3928	2.3479
21	4.3248	3.4668	3.0725	2.8401	2.6848	2.5727	2.4876	2.4205	2.3660	2.3210
22	4.3009	3.4434	3.0491	2.8167	2.6613	2.5491	2.4638	2.3965	2.3419	2.2967
23	4.2793	3.4221	3.0280	2.7955	2.6400	2.5277	2.4422	2.3748	2.3201	2.2747
24	4.2597	3.4028	3.0088	2.7763	2.6207	2.5082	2.4226	2.3551	2.3002	2.2547
25	4.2417	3.3852	2.9912	2.7587	2.6030	2.4904	2.4047	2.3371	2.2821	2.2365
26	4.2252	3.3690	2.9752	2.7426	2.5868	2.4741	2.3883	2.3205	2.2655	2.2197
27	4.2100	3.3541	2.9604	2.7278	2.5719	2.4591	2.3732	2.3053	2.2501	2.2043
28	4.1960	3.3404	2.9467	2.7141	2.5581	2.4453	2.3593	2.2913	2.2360	2.1900
29	4.1830	3.3277	2.9340	2.7014	2.5454	2.4324	2.3463	2.2783	2.2229	2.1768
30	4.1709	3.3158	2.9223	2.6896	2.5336	2.4205	2.3343	2.2662	2.2107	2.1646

Anexa 4. Valorile critice pentru coeficientul de corelație Pearson (r)

(sursa: http://www.psychology.ilstu.edu/psy138/tables.html)

Unilateral				
	p=0.05	***p=0.25***	***p=0.01***	***p=0.005***
Bilateral				
df	***p=0.10***	***p=0.05***	***p=0.02***	***p=0.01***
1	0,988	0,997	0,9995	0,9999
2	0,9	0,95	0,98	0,99
3	0,805	0,878	0,934	0,959
4	0,729	0,811	0,882	0,917
5	0,669	0,754	0,833	0,874
6	0,622	0,707	0,789	0,834
7	0,582	0,666	0,75	0,798
8	0,549	0,632	0,716	0,765

[22] Tabelul este aplicabil pentru maxim 11 grupuri ($df_{Between}$=10) și df_{Within} maxim=30. Tabele complete pentru F se găsesc în manualele recomandate în bibliografie.

Copyright © DEPARTAMENT ID 2008

Unilateral				
	p=0.05	***p=0.25***	***p=0.01***	***p=0.005***
Bilateral				
df	***p=0.10***	***p=0.05***	***p=0.02***	***p=0.01***
9	0,521	0,602	0,685	0,735
10	0,497	0,576	0,658	0,708
11	0,476	0,553	0,634	0,684
12	0,458	0,532	0,612	0,661
13	0,441	0,514	0,592	0,641
14	0,426	0,497	0,574	0,623
15	0,412	0,482	0,558	0,606
16	0,4	0,468	0,542	0,59
17	0,389	0,456	0,528	0,575
18	0,378	0,444	0,516	0,561
19	0,369	0,433	0,503	0,549
20	0,36	0,423	0,492	0,537
21	0,352	0,413	0,482	0,526
22	0,344	0,404	0,472	0,515
23	0,337	0,396	0,462	0,505
24	0,33	0,388	0,453	0,496
25	0,323	0,381	0,445	0,487
26	0,317	0,374	0,437	0,479
27	0,311	0,367	0,43	0,471
28	0,306	0,361	0,423	0,463
29	0,301	0,355	0,416	0,456
30	0,296	0,349	0,409	0,449
35	0,275	0,325	0,381	0,418
40	0,257	0,304	0,358	0,393
45	0,243	0,288	0,338	0,372
50	0,231	0,273	0,322	0,354
60	0,211	0,25	0,295	0,325
70	0,195	0,232	0,274	0,302
80	0,183	0,217	0,256	0,283
90	0,173	0,205	0,242	0,267
100	0,164	0,195	0,23	0,254

Copyright © DEPARTAMENT ID 2008

Anexa 5. Tabelul Fisher de transformare a valorilor *r* în scoruri Z
(Sursa: http://davidmlane.com/hyperstat/rtoz_table.html)

R	***Z***	***r***	***Z***	***r***	***Z***	***R***	***Z***
0.0000	*0.0000*	0.2600	*0.2661*	0.5200	*0.5763*	0.7800	*1.0454*
0.0100	*0.0100*	0.2700	*0.2769*	0.5300	*0.5901*	0.7900	*1.0714*
0.0200	*0.0200*	0.2800	*0.2877*	0.5400	*0.6042*	0.8000	*1.0986*
0.0300	*0.0300*	0.2900	*0.2986*	0.5500	*0.6184*	0.8100	*1.1270*
0.0400	*0.0400*	0.3000	*0.3095*	0.5600	*0.6328*	0.8200	*1.1568*
0.0500	*0.0500*	0.3100	*0.3205*	0.5700	*0.6475*	0.8300	*1.1881*
0.0600	*0.0601*	0.3200	*0.3316*	0.5800	*0.6625*	0.8400	*1.2212*
0.0700	*0.0701*	0.3300	*0.3428*	0.5900	*0.6777*	0.8500	*1.2562*
0.0800	*0.0802*	0.3400	*0.3541*	0.6000	*0.6931*	0.8600	*1.2933*
0.0900	*0.0902*	0.3500	*0.3654*	0.6100	*0.7089*	0.8700	*1.3331*
0.1000	*0.1003*	0.3600	*0.3769*	0.6200	*0.7250*	0.8800	*1.3758*
0.1100	*0.1104*	0.3700	*0.3884*	0.6300	*0.7414*	0.8900	*1.4219*
0.1200	*0.1206*	0.3800	*0.4001*	0.6400	*0.7582*	0.9000	*1.4722*
0.1300	*0.1307*	0.3900	*0.4118*	0.6500	*0.7753*	0.9100	*1.5275*
0.1400	*0.1409*	0.4000	*0.4236*	0.6600	*0.7928*	0.9200	*1.5890*
0.1500	*0.1511*	0.4100	*0.4356*	0.6700	*0.8107*	0.9300	*1.6584*
0.1600	*0.1614*	0.4200	*0.4477*	0.6800	*0.8291*	0.9400	*1.7380*
0.1700	*0.1717*	0.4300	*0.4599*	0.6900	*0.8480*	0.9500	*1.8318*
0.1800	*0.1820*	0.4400	*0.4722*	0.7000	*0.8673*	0.9600	*1.9459*
0.1900	*0.1923*	0.4500	*0.4847*	0.7100	*0.8872*	0.9700	*2.0923*
0.2000	*0.2027*	0.4600	*0.4973*	0.7200	*0.9076*	0.9800	*2.2976*
0.2100	*0.2132*	0.4700	*0.5101*	0.7300	*0.9287*	0.9900	*2.6467*
0.2200	*0.2237*	0.4800	*0.5230*	0.7400	*0.9505*		
0.2300	*0.2342*	0.4900	*0.5361*	0.7500	*0.9730*		
0.2400	*0.2448*	0.5000	*0.5493*	0.7600	*0.9962*		
0.2500	*0.2554*	0.5100	*0.5627*	0.7700	*1.0203*		

Copyright © DEPARTAMENT ID 2008

UNIVERSITATEA DIN BUCUREŞTI

FACULTATEA DE PSIHOLOGIE ŞI ŞTIINȚELE EDUCAȚIEI

DEPARTAMENTUL DE ÎNVĂȚĂMÂNT LA DISTANȚĂ

STATISTICĂ PSIHOLOGICĂ ŞI PRELUCRAREA COMPUTERIZATĂ A DATELOR

(Modulul III)

Teste statistice neparametrice

Conf. univ. dr. Marian Popa
Email: popamarian@gmail.com

Universitatea din Bucureşti
Editura CREDIS
2008

Copyright © DEPARTAMENT ID 2008

CUPRINS

TESTE NEPARAMETRICE PENTRU DATE NOMINALE....28
Obiectivele unității de învățare....28
Noțiuni introductive....28
Distribuția binomială....29
Testul z pentru proporția unui eșantion....33
Testul *z* pentru proporția unui eșantion în raport cu populația....33
Sarcina de lucru nr. 3. 1....34
Testul z pentru două proporții independente....34
Sarcina de lucru nr. 3. 2....35
Testul semnului pentru eșantioane dependente....36
Sarcina de lucru nr. 3. 3....37
Distribuția multinomială....37
Testul chi-pătrat....38
Tabelul de corespondență (contingență)....38
Fundamentarea testului statistic....39
Testul chi-pătrat pentru gradul de corespondență (*goodness of fit*)....40
Sarcina de lucru nr. 3. 4....42
Chi-pătrat - testul asocierii (independence chi-square)....42
Marimea efectului pentru testul *chi pătrat* al asocierii....44
Sarcină de lucru nr. 3. 5....46
Testul exact Fisher....46
Rezumatul unității de învățare....47
Răspunsuri şi comentarii la problemele de evaluare....48
Lucrarea de evaluare nr. 3.1 şi modul de cotare....50
TESTE STATISTICE NEPARAMETRICE PENTRU DATE ORDINALE....51
Obiectivele unității de învățare şi informații introductive....51
Introducere....51
Testul Mann-Whitney (U) pentru două eșantioane independente....51
Sarcina de lucru nr. 3.6....54
Testul Kruskall-Wallis pentru mai mult de două eșantioane independente....54
Sarcina de lucru nr. 3.7....56
Testul Wilcoxon pentru două eșantioane perechi....56
Sarcina de lucru 3.8....57
Testul Friedman pentru măsurări repetate....58
Sarcina de lucru nr. 3.9....59
Coeficientul de corelație pentru date ordinale (Spearman)....59
Sarcina de lucru 3.10....61
Rezumatul unității de învățare....62
Răspunsuri la sarcinile de lucru....63
Lucrarea de evaluare nr. 3.2 şi modul de cotare....66
Bibliografie minimală....66
ANEXE – TABELE STATISTICE....67
Anexa 6. Valorile critice pentru distribuția *chi-pătrat*....67
Anexa 7. Tabelul valorilor critice pentru testul Mann-Whitney U....68
Anexa 8. Valorile critice pentru testul Wilcoxon....69
Anexa 9. Valorile critice pentru testul de corelație a rangurilor (Spearman)....70

Copyright © DEPARTAMENT ID 2008

TESTE NEPARAMETRICE PENTRU DATE NOMINALE

Obiectivele unității de învățare

Parcurgerea acestei unități va permite studenților:

- ✓ să definească proprietățile distribuției binomiale
- ✓ să testeze semnificația procentului unui eşantion în raport cu populația
- ✓ să testeze semnificația diferenței dintre două proporții
- ✓ să testeze diferența dintre două eşantioane dependente evaluate pe scală nominală
- ✓ să definească proprietățile distribuției multinomiale
- ✓ să testeze semnificația asocierii (corespondenței) dintre variabile nominale

Noțiuni introductive

Testele neparametrice prezintă, în raport cu cele parametrice, o serie de avantaje, dar şi dezavantaje.

Principalele **avantaje** sunt:

- Se pot utiliza pe scale ale căror calități de măsurare sunt „slabe" (ordinale, nominale).
- Pot fi utilizate în cazul variabilelor afectate de valori extreme care nu pot fi eliminate.
- Utilizarea lor nu presupune condiții la fel de restrictive ca testele parametrice (normalitatea distribuției, omogenitatea varianței, etc.).
- Pentru anumite proceduri, calculele sunt relativ simple şi uşor de efectuat, chiar şi fără utilizarea tehnicii de calcul.
- Conceptele şi metodele statisticii neparametrice sunt uşor de înțeles.

Printre **dezavantajele** testelor neparametrice, sunt de menționat:

- Se bazează pe măsurări pe scale nominale şi ordinale, care sunt, prin natura lor, măsurări mai puțin precise decât cele pe scale cantitative (de interval sau de raport).
- Au o „putere" mai redusă decât testele parametrice de a proba că ipoteza cercetării este adevărată.
- Tind sa fie utilizate, datorită relativei lor simplități, şi în situații în care se pot utiliza teste parametrice. Este important să reținem faptul că, atunci când sunt întrunite condițiile pentru aplicarea unui test parametric, nu este recomandabilă transformarea variabilei şi utilizarea unui test neparametric.
- Deşi se bazează pe calcule elementare, adesea acestea pot fi destul de complexe şi de laborioase.
- Principiul care stă la baza evaluării mărimii efectului pentru testele parametrice (proporția explicată a varianței) nu este uşor de aplicat în cazul testelor neparametrice. Ca urmare,

Copyright © DEPARTAMENT ID 2008

pentru multe dintre testele neparametrice nu se poate calcula un indicie de mărime a efectului.

Ca o concluzie generală, utilizarea testelor neparametrice nu poate fi evitată dacă variabila dependentă este una de tip nominal sau ordinal. Dacă, însă, este măsurată pe o scală de interval/raport, se pune problema de a alege între un test parametric şi unul neparametric. În acest caz, criteriul principal de decizie este normalitatea distribuţiei la nivelul populaţiei. În principiu, teorema limitei centrale oferă suportul teoretic al asumării acestei condiţii pentru eşantioane „suficient de mari". Din păcate, nu avem nici un criteriu sigur de verificare a acestei condiţii. Din acest motiv există o anumită dispută în legătură cu justificarea utilizării testelor parametrice în anumite cazuri. Dacă eşantioanele care se apropie sau depăşesc 100 de valori (subiecţi) permit asumarea cu încredere a condiţiei de normalitate, eşantioanele de mărimi medii (20-40 de subiecţi) sunt considerate mai puţin sigure. Simulările pe calculator au arătat că există teste parametrice mai puţin vulnerabile la violarea condiţiei de normalitate (testele *t*, de exemplu), dar şi altele care devin nesigure în această situaţie (de ex., testul F pentru omogenitatea varianţei). Fără a încerca tranşarea disputei, putem reţine că, mai ales pentru eşantioanele mici, atunci când avem motive să ne îndoim de normalitatea distribuţiei la nivelul populaţiei, vor fi preferate testele neparametrice.

Distribuţia binomială

Atunci când măsurăm o caracteristică pe o scală de tip cantitativ, obţinem o valoare care descrie „mărimea" acelei caracteristici. Uneori, însă, nu facem decât să observăm măsura în care acea caracteristică este prezentă într-un anumit context. De exemplu, observăm caracteristica de gen (masculin, feminin) a copiilor la naştere, „prezenţa"⁄„absenţa" efectului unei metode psihoterapeutice sau caracterul „corect"⁄„greşit" al răspunsului la o serie de întrebări. În acest toate aceste cazuri naşterea unui băiat (sau unei fete), „prezenţa efectului", „răspunsul corect" sunt denumite „evenimente" despre care putem doar să consemnăm frecvenţa cu care apar într-o anumită serie de „încercări" (naşteri, subiecţi trataţi cu metoda respectivă, listă de întrebări).

Distribuţia evenimentelor de tip dihotomic, fiecare având o anumită probabilitate de apariţie, constantă de la o încercare la alta, se numeşte distribuţie binomială[1]. Caracteristicile distribuţiei binomiale diferă în funcţie de numărul „încercărilor" (N) şi de probabilitatea de apariţie a „evenimentului" (P), văzută ca şansă teoretică de apariţie a evenimentului, în raport cu toate evenimentele posibile. De exemplu, la aruncarea unei monede, o singură dată (N=1), şansa (probabilitatea) teoretică de apariţie a „stemei" este P=1/2=0.5. Aceeaşi probabilitate caracterizează şi evenimentul „răspuns corect", dacă răspundem la întâmplare la o întrebare cu două variante de răspuns, dintre care una este corectă iar alta greşită.

Să transpunem această problemă într-o situaţie cu relevanţă practică. Să ne imaginăm că am construit un chestionar de cunoştinţe de statistică, compus din întrebări cu două variante de răspuns, una corectă şi una eronată. În faţa rezultatelor, este firesc să ne întrebăm dacă studenţii au răspuns utilizându-şi cunoştinţele, sau la întâmplare, încercându-şi norocul. Dacă la un chestionar cu patru întrebări un student dă patru răspunsuri corecte, sunt ele un indiciu suficient că rezultatul reflectă cunoştinţele de statistică şi nu norocul de a fi ghicit de fiecare dată răspunsul corect?

Pentru a încerca să rezolvăm aceasta dilemă, să zicem că ne adresăm unui alt student pentru a răspunde absolut la întâmplare. Ca să fim siguri că răspunsurile acestuia nu sunt „alterate" de cunoştinţele sale de statistică, îi cerem să aleagă răspunsul fără a vedea întrebările, dând de patru ori cu banul, pentru a indica răspunsul la fiecare întrebare. În acest caz, răspunsurile corecte decurg numai prin jocul hazardului, adică sunt „evenimente aleatoare". Acestea se definesc ca raport între evenimentul aşteptat şi numărul evenimentelor posibile. Existând doar două variante de răspuns, probabilitatea teoretică de a răspunde corect la o întrebare este de 0.5. Probabilitatea de a răspunde corect la toate cele patru întrebări se calculează ca produs al probabilităţii fiecărui element al secvenţei de patru întrebări (regula multiplicării probabilităţii evenimentelor dihotomice):

[1] Sau „distribuţie Bernoulli" , după numele matematicianului elveţian Jakob Bernoulli (1654-1705).

Copyright © DEPARTAMENT ID 2008

0.5*0.5*0.5*0.5*=0.0625

Constatăm astfel că, răspunzând absolut la întâmplare, probabilitatea de ghici toate răspunsurile corecte este de 0.0625. Nu este o probabilitate foarte mare, dar este, totuşi, mai mare decât nivelul alfa minim de 0.05, cu care ne-am obişnuit deja. Ca urmare, suntem nevoiți să acceptăm că cele patru răspunsuri corecte sunt mai degrabă un rezultat al unor alegeri întâmplătoare decât al cunoştințelor de statistică. Concluzia ar fi că, dacă dorim să păstrăm tipul de întrebări cu două variante de răspuns, atunci va trebui cel puțin să mărim numărul întrebărilor. Astfel, să zicem, vom ajunge în situația de a ne pune problema dacă putem avea încredere într-un rezultat de 8 răspunsuri corecte din 10 întrebări.

Dar pe măsură ce numărul alegerilor binare (cu două variante posibile de răspuns) creşte, calcularea probabilității răspunsurilor întâmplătoare se complică. Din acest motiv devine necesară o anumită formalizare a situației. Distribuția probabilităților pentru evenimente dihotomice aleatoare alcătuieşte distribuția binomială.

Ea prezintă interes ca distribuție de nul pentru cazuri ca cele din exemplul de mai sus. Având un eveniment cu doar două variante, fiecare cu şansă egală, (de ex., masculin/feminin, corect/greşit), vom nota cu P probabilitatea uneia dintre variante şi cu Q probabilitatea variantei complementare. Întotdeauna P+Q=1, ceea ce face posibil să-l descriem Q sub forma Q=1-P.

O distribuție binomială se obține pe baza unei secvențe de predicții de tip dihotomic, independente între ele, pentru care valoarea lui P şi Q nu se modifică de la o predicție la alta. O astfel de selecție este şi cea făcută de studentul care a indicat răspunsurile corecte, dând cu banul la cele patru întrebări de statistică. Numărul total de predicții (în exemplul nostru, 4) este simbolizat cu N. Dată fiind relația dintre P şi Q, este suficient să analizăm predicția pentru unul dintre cele două evenimente posibile, să zicem pentru răspunsurile „corecte", deoarece probabilitățile pentru evenimentul complementar (răspunsuri greşite) sunt absolut simetrice. Distribuția binomială depinde, în acelaşi timp, de valoare lui P şi a lui N.

Să analizăm variația predicțiilor pentru cele patru întrebări de statistică. Toate combinațiile posibile între răspunsurile corecte (C) şi eronate (E) se pot afla prin listarea combinațiilor şi permutările posibile (2*2*2*2=16) pentru cele patru întrebări:

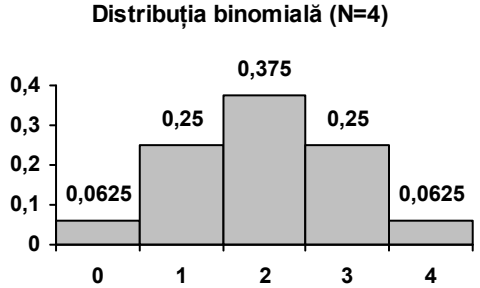
CCCC	CECC	ECCC	EECC
CCCE	CECE	ECCE	EECE
CCEC	CEEC	ECEC	EEEC
CCEE	CEEE	ECEE	EEEE

Dacă analizăm toate cele 16 combinații posibile, vom observa că avem următoarea distribuție probabilă pentru răspunsurile corecte:

Nr. răsp. corecte	0	1	2	3	4
Frecvența	1	4	6	4	1
P(C)*	1/16=0.0625	4/16=0.25	6/16=0.375	4/16=0.25	1/16=0.0625

P(C) =Probabilitatea de apariție a răspunsului corect

Transpuse grafic, probabilitățile corespunzătoare pentru frecvențele de răspuns corect se prezintă ca în imaginea următoare:



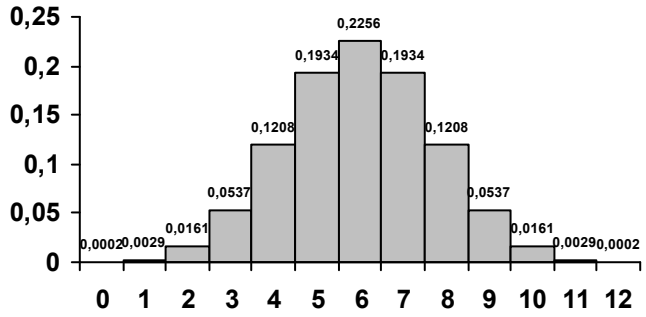
Cu alte cuvinte, în cazul alegerii întâmplătoare a unui răspuns din două posibile, pentru patru întrebări, probabilitatea niciunui răspuns corect este egală cu aceea pentru patru răspunsuri corecte (0.0625).

Copyright © DEPARTAMENT ID 2008

Cea mai mare probabilitate o are situația de a rezulta două răspunsuri corecte (0.375), în timp ce probabilitatea de a ghici 1 sau trei răspunsuri corecte este de 0.25. Nu putem să nu observăm, de asemenea, forma simetrică a distribuției[2]. În conformitate cu teorema Moivre-Laplace, distribuția binomială a apariției evenimentelor echiprobabile (P=Q=0.5) într-o serie de *n* de observații independente, urmează forma distribuției normale. Sau, mai exact, după standardizarea probabilităților acestea corespund valorilor de sub curba normală.

Dar ce s-ar întâmpla dacă, în loc de 4 întrebări, chestionarul nostru de statistică ar avea 12 întrebări? Distribuția binomială pentru N=12 este cea din graficul de mai jos:

Distribuția binomiala (N=12)



Se observă creşterea corespunzătoare a numărului variantelor posibile şi, în acelaşi timp, devine mai evidentă tendința distribuției de a semăna cu una normală. În mod firesc, această tendință se accentuează pe măsură ce numărul secvențelor de predicție creşte.

Dar sunt şi situații în care P şi Q nu sunt egale. De exemplu, dacă variantele de răspuns la fiecare întrebare a chestionarului de statistică sunt în număr de patru, dintre care numai una este corectă, atunci probabilitatea răspunsului corect (P) este ¼=0.25. În acest caz distribuția binomială nu este simetrică la valori mici ale lui N, dar tinde să devină simetrică pe măsură ce N creşte. Nu există un răspuns exact cu privire la valoarea lui N pentru care distribuția binomială este aproximată suficient de bine de cea normală. În general, se acceptă faptul că pentru P=0.5 N nu trebuie să fie mai mare de 20-25, în timp ce pentru P apropiat de 0 sau 1 se impune o valoare pentru N de cel puțin 100.

Din cele spuse rezultă că se poate lua în considerare aproximarea distribuției binomiale cu o distribuție normală. Aceasta înseamnă că putem exprima valorile *z* în termeni de N, P şi Q. Formula originală pentru ***z*** ne amintim că este:

$$z = \frac{X - \mu}{\sigma}$$

din care, prin substituire, se construieşte formula pentru ***z*** binomial:

$$z = \frac{X - N * P}{\sqrt{N * P * Q}} \quad \text{(formula 4.1)}$$

Această formulă poate fi utilizată pentru a afla câtă încredere putem avea în cazul în care am obține 8 răspunsuri corecte la un chestionar cu 10 întrebări dihotomice:

$$z = \frac{8 - 10 * 0.5}{\sqrt{10 * 0.5 * 0.5}} = \frac{8 - 5}{\sqrt{2.5}} = \frac{3}{1.581} = 1.897$$

Nivelul probabilității de sub curba normală *z*, pentru valori ale lui *z* egale sau mai mari de 1.897 este 0.0294. Aceasta înseamnă că putem respinge ipoteza de nul şi să admitem că studentul nu a răspuns la

[2] Distribuția binomială a fost descrisă de Abraham De Moivre în lucrarea *„Approximatio ad Summam Terminorum Binomii in Seriem Expansi", publicată în* 1733. Același autor a publicat şi un manual pentru jucătorii de noroc, în care descrie principiile aritmetice pentru strategiile şi probabilitățile de câştig.

Copyright © DEPARTAMENT ID 2008

întâmplare. Vom observa însă, că putem accepta această concluzie numai dacă, anterior calculelor, am ales o decizie de tip unilateral, deoarece pentru o decizie bilaterală ar fi fost necesar un nivel minim p=0.025. Oricum, constatarea cea mai importantă în acest caz este aceea că utilizarea întrebărilor cu răspuns dihotomic nu este recomandabilă, din cauza şansei prea mari de se obține un număr relativ ridicat de răspunsuri corecte prin alegeri întâmplătoare.

Să schimbăm puțin datele problemei şi să punem la fiecare întrebare nu două, ci patru variante de răspuns, dintre care numai una este corectă. În acest caz, P=1/4=0.25 iar Q=3/4=0.75. Considerând un chestionar format tot din 10 întrebări, cu 8 răspunsuri corecte, şi utilizând formula 4.1, valoarea testului de semnificație este:

$$z = \frac{8 - 10 * 0.25}{\sqrt{10 * 0.25 * 0.75}} = \frac{8 - 2.5}{\sqrt{1.875}} = \frac{5.5}{1.369} = 4.01$$

În aceste condiții este evident că ipoteza de nul se respinge, iar ipoteza că răspunsurile se bazează mai mult pe cunoştințe decât pe hazard se acceptă. Fără să reluăm calculele, putem să ne dăm seama că am obține o valoare semnificativă chiar şi pentru un număr mai mic de răspunsuri corecte. Desigur, acesta este un exemplu didactic, în practică nefiind utilizate chestionare de cunoştințe cu un număr atât de mic de întrebări.

Copyright © DEPARTAMENT ID 2008

Testul z pentru proporția unui eşantion

Testul *z* pentru proporția unui eşantion în raport cu populația

Odată ce am găsit o modalitate de elaborare a distribuției de nul pentru evenimente de tip binomial, se pot elabora diverse teste de inferență statistică. Unul dintre acestea este testul ***z*** pentru proporții, care este echivalentul pentru date nominale al testului ***z*** parametric pentru un singur eşantion.

Să ne imaginăm situația în care descoperim că, pe un eşantion aleator de 100 de subiecți dintr-o anumită comunitate, procentul stângacilor este de 20%, în timp ce studiile la nivelul populației generale indică un procent de stângaci de numai 15% . În acest caz ne putem pune întrebarea dacă la nivelul acelei comunități există o „anomalie" a lateralității.

Pentru a putea utiliza formula 4.1 pentru testarea directă a proporțiilor, o supunem unei transformări convenabile, prin împărțirea simultană a numărătorului şi numitorului cu N. Ca urmare, obținem următoare formulă:

$$z = \frac{p - P}{\sqrt{\frac{PQ}{N}}} \qquad \text{(formula 4.2)}$$

unde: *p* (mic) este probabilitatea măsurată a evenimentului cercetat,
P (mare) este probabilitatea aceluiaşi eveniment la nivelul populației,
Q este probabilitatea complementară a lui P,
N este volumul eşantionului.

Pentru cazul nostru, valoarea testului *z* pentru proporții se obține astfel:

$$z = \frac{0.20 - 0.15}{\sqrt{\frac{0.15 * 0.85}{100}}} = \frac{0.05}{\sqrt{\frac{0.127}{100}}} = \frac{0.05}{0.035} = 1.42$$

Nivelul lui *p* pentru *z*=1.42 pe curba normală este de 0.0778, valoare care obligă la acceptarea ipotezei de nul. Cu alte cuvinte, proporția stângacilor în comunitatea cercetată nu depăşeşte semnificativ proporția la nivelul populației generale.

Testul ***z*** pentru proporții implică testarea semnificației unui procent observat în raport procentul populației (atunci când este cunoscut), pentru evenimente de tip dihotomic. De exemplu, se poate răspunde la întrebarea dacă un procent 55% de nou născuți băieți este neobişnuit de mare, ştiind care este procentul general al noilor născuți băieți.

Pentru situațiile în care evenimentele cercetate nu sunt de tip dihotomic, se aplică alte teste statistice, despre care vom vorbi mai târziu.

.

Copyright © DEPARTAMENT ID 2008

Sarcina de lucru nr. 3. 1

(1) Un profesor constată că unul dintre studenți a dat 58 de răspunsuri corecte la un examen cu 100 de întrebări, fiecare cu două variante de răspuns (ADEVĂRAT-FALS). A răspuns studentul la întâmplare, sau rezultatul obținut decurge dintr-o anumită cunoaştere a materiei? (alfa=0.05, bilateral)

(2) Presupunând că 85% din populație este dreptace (Q) şi că 15% este stângace (P):
   a. Dacă 27 din cei 120 de copii dintr-o şcoală de artă sunt stângaci, care este scorul z pentru testarea ipotezei?
   b. Pe baza scorului z de la punctul a, putem concluziona că frecvența stângacilor printre copiii cu aptitudini artistice este mai mare decât la nivelul populației? (alfa=0.05, bilateral)

Verificați răspunsurile corecte

Testul z pentru două proporții independente

Să ne întoarcem la exemplul de mai sus, cu privire la proporția stângacilor, şi să îl privim din altă perspectivă. Un studiu pe două eşantioane din două țări diferite conduce la constatarea că proporția (p1=0.15) stângacilor a eşantionului (n1=100) dintr-o țară este diferită de proporția (p2=0.25) stângacilor din eşantionul corespunzător celeilalte țări (n2=90). Este firesc să ne punem întrebarea dacă există într-adevăr o diferența dintre proporția stângacilor din cele două țări (pe care o vom nota cu litere mari: P1 respectiv P2) sau dacă, dimpotrivă, diferențele constatate sunt doar expresia variabilității de eşantionare.

În acest caz:

- ipoteza cercetării susține că proporțiile la nivelul populațiilor sunt diferite (P1≠P2)
- ipoteza de nul susține că proporțiile celor două populații sunt identice (P1=P2) şi, deci, că diferența lor este 0 (P1-P2=0)

În exemplul nostru, P1 şi P2 reprezintă probabilitățile unui eveniment aleator de tip binomial, în care evenimentul complementar (Q1, respectiv Q2) este caracteristica de a fi „dreptaci" (vom ignora acum faptul că pot exista şi „ambidextri").

Distribuția ipotezei de nul pentru diferențele dintre cele două proporții este aproximată de distribuția normală z. Testul statistic va urma modelul testului pentru diferența dintre mediile a două eşantioane independente:

$$z = \frac{(p_1 - p_2) - (P_1 - P_2)}{\sigma_{(p_1 - p_2)}} \qquad \text{(formula 4.3)}$$

unde:

p_1 şi p_2 sunt proporțiile evenimentului la nivelul eşantioanelor

P_1 şi P_2 sunt proporțiile evenimentului la nivelul populației

$\sigma_{(p1-p2)}$ este eroarea standard a distribuției de eşantionare

Având în vedere ipoteza de nul (P_1-P_2=0), rezultă că la numitor se va păstra doar diferența dintre proporțiile eşantioanelor (p_1-p_2).

Copyright © DEPARTAMENT ID 2008

La rândul ei, eroarea standard de eşantionare a diferenţei proporţiilor se calculează astfel:

$$\sigma_{(p_1-p_2)} = \sqrt{\frac{p_1 * q_1}{n_1} + \frac{p_2 * q_2}{n_2}} \quad \text{(formula 4.4)}$$

unde:
q1 şi q2 sunt proporţiile complementare ale lui p_1, respectiv p_2 (q_1=1-p_1, respectiv q_2=1-p_2)
n_1 şi n_2 sunt volumele celor două eşantioane

Ca urmare, formula pentru testul diferenţei dintre proporţiile a două eşantioane independente devine:

$$z = \frac{p_1 - p_2}{\sqrt{\frac{p_1 * q_1}{n_1} + \frac{p_2 * q_2}{n_2}}} \quad \text{(formula 4.5)}$$

Această formulă este adecvată atunci când eşantioanele sunt suficient de mari (>30). În caz contrar, numărătorul formulei suportă o corecţie, după cum urmează:

$$z = \frac{\left(p_1 - \frac{1}{2 * n_1}\right) - \left(p_2 - \frac{1}{2 * n_2}\right)}{\sqrt{\frac{p_1 * q_1}{n_1} + \frac{p_2 * q_2}{n_2}}} \quad \text{(formula 4.6)}$$

Pentru exemplul nostru, vom utiliza formula 4.5

$$z = \frac{0.15 - 0.25}{\sqrt{\frac{0.15 * 0.85}{100} + \frac{0.25 * 0.75}{90}}} = \frac{-0.10}{\sqrt{0.001 + 0.002}} = \frac{-0.10}{0.054} = -1.85$$

Dacă ne-am propus un test bilateral la un nivel alfa=0.05 (pentru care z critic pe curba normală este egal cu 1.96), atunci va trebui să acceptăm ipoteza de nul şi să concluzionăm că nu se confirmă existenţa unei diferenţe semnificative între proporţia stângacilor din cele două comunităţi.

Sarcina de lucru nr. 3. 2

Două grupuri de subiecţi, fiecare compus din 30 de persoane, participă la un experiment în care este studiat efectul stresului temporal asupra performanţei de rezolvare de probleme. Primul grup are un termen limită iar celalalt, nu are un termen limită. Rezultatele cercetării arată că 25% dintre subiecţii grupului care a lucrat în criză de timp au rezolvat problema, în timp ce pentru grupul fără criză de timp, procentul rezolvărilor corecte este de 60%. Se poate afirma că stresul temporal reduce performanţa în rezolvarea de probleme? (alfa=0.05, bilateral)

Verifica□i corectitudinea răspunsului

Copyright © DEPARTAMENT ID 2008

Testul semnului pentru eşantioane dependente

Ne amintim că unul dintre modelele uzuale de cercetare în psihologie este cel care se bazează pe eşantioane perechi (corelate sau dependente), în care este evaluată o anumită variabila de două ori pentru aceiaşi subiecţi (sau perechi de subiecţi). Dacă rezultatul măsurării este exprimat pe o scală de interval/raport, atunci diferenţa dintre cele două momente (situaţii) se verifică cu ajutorul testului *t* pentru eşantioane dependente. Ce ne facem, însă, dacă nu dispunem de posibilitatea unei măsurări la nivel cantitativ şi suntem nevoiţi să observăm doar sensul variaţiei de la un moment la altul? Soluţia acestei probleme a fost găsită în anul 1710 de John Arbuthnot[3], medicul personal al reginei Anna a Angliei, primul care a utilizat testul semnului în analiza retrospectivă, pe o perioadă de 82 de ani, a raportului naşterilor de băieţi şi fete (13/12), înregistrate la primăria Londrei.

Să ne imaginăm următoarea situaţie de cercetare: un psiholog clinician aplică o metodă de reducere a manifestărilor de tip fobic la un grup de 8 de subiecţi. După un număr de şedinţe el doreşte să afle dacă metoda lui este eficientă şi îi întreabă pe cei 8 subiecţi dacă se simt mai bine decât la începutul tratamentului. Răspunsurile arată că 6 dintre ei afirmă că se simt mai bine iar 2, că nu simt nici o modificare (să admitem că nimeni nu răspuns că se simte mai rău).

În acest caz ipoteza cercetării susţine că metoda are efect, ceea ce înseamnă că procentul de ameliorare este semnificativ mai mare decât cel al absenţei oricărui efect al terapiei. Ipoteza de nul este opusul ei, fapt care se exprimă prin echivalenţa celor două evenimente posibile (eficienţa/ineficienţa terapiei) şi se formalizează ca P=Q=0.5.

Având o probabilitate de 6/8=0.75 pentru evenimentul „ameliorare”, se poate afirma că acesta este semnificativ diferit de cel al ipotezei de nul (0.5)?

Pentru a verifica ipoteza, se utilizează formula 4.1:

$$z = \frac{X - N * P}{\sqrt{N * P * Q}}$$

Deşi, principial, este corectă, se impune o anumită corecţie a acestei formule, corecţie utilă mai ales pentru valori mici ale lui N. Dacă privim graficele distribuţiilor binomiale prezentate anterior vom observa că, spre deosebire de curba normală *z*, acestea au un caracter „discontinuu”, cu treceri în „trepte” la o valoare la alta. Din acest motiv se recomandă aplicarea unei „corecţii de continuitate”, prin scăderea valorii 0.5 din valoarea numărătorului, luată în sens absolut. Formula definitivă devine astfel:

$$z = \frac{|X - N * P| - 0.5}{\sqrt{N * P * Q}} \quad \text{(formula 4.7)}$$

Mai departe, nu ne rămâne decât să înlocuim valorile şi sa facem calculele pentru studiul nostru:

$$z = \frac{|6 - 8 * 0.75| - 0.5}{\sqrt{8 * 0.75 * 0.25}} = \frac{-0.5}{1.22} = -0.40$$

Mai departe, căutăm valoarea lui *p* corespunzătoare pentru z=-0.40 în tabelul distribuţiei normale *z*, unde găsim *p*=0.844. Dat fiind faptul că valoarea lui *p* este mai mare decât 0.05, suntem nevoiţi să acceptăm ipoteza de nul şi să conchidem că, cel puţin până în acel moment, terapia antifobică nu are un efect semnificativ statistic pe lotul aflat în tratament. Desigur, rezultatul nu trebuie să fie considerat neapărat ca descurajant de către terapeut. Faptul că lotul investigat este atât de redus conduce în mod inevitabil la nevoia unor valori foarte ridicate ale testului statistic pentru atingerea pragului de semnificaţie. În cazul nostru rezultatul poate fi considerat încurajator dacă, să zicem, evaluarea eficienţei s-a făcut după un număr relativ mic de şedinţe de terapie. Continuarea lor şi refacerea testului ar putea conduce la o altă concluzie.

[3] Arbuthnot, John. (1710), "An Argument for Divine Providence, Taken From the Constant Regularity Observed in the Births of Both Sexes," *Philosophical Transactions*, 27, 186-190.

Copyright © DEPARTAMENT ID 2008

Testul semnului (denumit astfel pentru că ia în considerare doar sensul variației, nu şi valoarea ei) este utilizabil ca substitut al testului *t* pentru eşantioane dependente în cazul datelor măsurate pe scală nominală dihotomică.

Sarcina de lucru nr. 3. 3

(1) Într-un experiment de psihologie socială, participanților (N=35) li s-a cerut să citească un pasaj care conține o dilemă a cărei soluție este apoi descrisă. Subiecților li s-a cerut apoi să indice nivelul de susținere pe care îl manifestă față de soluția propusă, pe o scală de la 1 („foarte puternic împotrivă"), la 7 („foarte mult în favoarea soluției"). În continuare, au loc discuții pe tema soluției, după care subiecților li se cere să îşi exprime din nou gradul de susținere. Rezultatele indică faptul că, în urma discuțiilor, un număr de 30 de subiecți şi-au modificat gradul de susținere inițială a soluției. Problema cercetării este dacă discuția în grup are un efect asupra opiniei participanților.

(2) Şase studenți de la facultatea de arte plastice au fost puşi să picteze două tablouri, pe o temă imaginară. Într-un caz au lucrat în condiții de liniște, în cel de-al doilea caz au avut un fond sonor de muzică clasică. Lucrările lor au fost evaluate de un profesor, care a apreciat că 5 dintre studenți au pictat mai creativ în condiții de muzică decât în condiții de liniște. Se poate concluziona că muzica clasică favorizează creativitatea artistică, pentru alfa=0.05, bilateral? (utilizați Tabelul 3.1 din Anexe)

Verifică răspunsul corect

Distribuția multinomială

Să presupunem că populația absolvenților de liceu se împarte în patru categorii: „*teoretic-umanist"*, „*teoretic-real*", „*artistic"*. Probabilitățile aferente fiecărui tip de liceu sunt, respectiv, P, Q şi R. Într-o asemenea situație P+Q+R=1. De asemenea, putem scrie probabilitățile pentru fiecare „eveniment" după modelul: Q=1-P-R. Să luăm în considerare situația în care toate liceele au acelaşi număr de absolvenți, deci P=Q=R=1/3=0.33. Alegerea unor ponderi diferite, aşa cum este şi cazul în realitate, nu ar schimba datele raționamentului care urmează dar l-ar face mai puțin evident. Mai departe, să ne imaginăm că analizăm tipul de liceu absolvit de studenții unei facultăți de psihologie şi constatăm că din 100 de studenți 60 sunt absolvenți de liceu „teoretic-umanist", 30 au absolvit un liceu cu profil „artistic" şi 10, unul cu profil „teoretic-real". Ponderea studenților la facultatea respectivă este, evident, diferită de ponderea din cadrul populației de absolvenți. Se poate afirma că absolvenții de profil „umanist" şi „artistic" preferă psihologia mai mult decât care au absolvit un profil „real"? Sau, într-o formulare mai largă, se poate afirma că există o relație între tipul de liceu absolvit şi preferința pentru psihologie ca specialitate universitară?

Datele din exemplul de mai sus nu mai pot fi analizate prin prisma distribuției binomiale deoarece implică mai mult decât două „evenimente" posibile. De aceea, distribuția acestora se numeşte „distribuție multinomială". Desigur, procedura de calcul pentru acest caz ar putea urma modelul celei binomiale dar, din cauza complexității ei, s-a apelat la o soluție mai simplă. Această

Copyright © DEPARTAMENT ID 2008

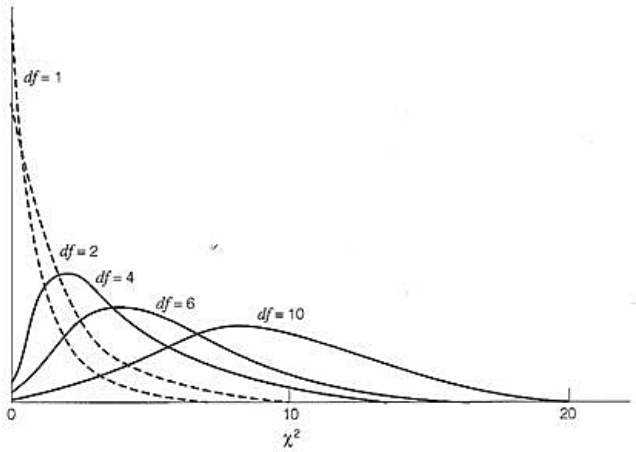
soluție este fundamentată pe o aproximare derivată din formula binomială a lui z, care este pur si simplu ridicată la pătrat, devenind:

$$z^2 = \frac{(X - N*P)^2}{N*P*Q} \quad \text{(formula 4.8)}$$

Dacă, înainte de ridicarea la pătrat, ***z*** urmează o distribuție normală, după ridicarea la pătrat z urmează un alt tip de distribuție, numită „*chi-pătrat*" şi simbolizată cu litera grecească χ cu indicele de ridicare la pătrat (χ^2). Fără a intra în amănunte, vom preciza că distribuția χ^2 prezintă următoarele caracteristici:

- este, la fel ca distribuția normală, o familie de distribuții;
- are formă asimetrică;
- are originea în zero (din cauza ridicării la pătrat);
- are o formă dependentă de numărul de grade de libertate.

Imaginea de mai jos prezintă două distribuții chi-pătrat, pentru 4, respectiv, 10 grade de libertate.



Înainte de a merge mai departe, să analizăm puțin, datele sugerate de exemplul de mai sus. Aşa cum am spus, ponderea studenților la facultatea de psihologie este, în funcție de tipul de liceu absolvit, de, respectiv, 60, 30, 10. Aceste valori se numesc „frecvențe observate" sau „frecvențe calculate" (notate cu f_o de la *Observed*), fiind cele consemnate în cadrul studiului. La rândul lor, frecvențele corespunzătoare distribuției de nul, cele care se referă, în acest caz, la ponderea absolvenților de liceu în general, se numesc „frecvențe aşteptate" (notate cu f_e de la *Expected*). Raportul dintre aceste două categorii de frecvențe se află la baza testului de semnificație statistică numit „*chi-pătrat*", despre care vorbim mai departe.

Testul chi-pătrat

Tabelul de corespondență (contingență)

Copyright © DEPARTAMENT ID 2008

Înainte de a trece la testul propriu-zis, este util să aruncăm o privirea asupra modului de organizare a datelor pentru o situație similară exemplului de mai sus. În acest scop, putem să ne permitem o lărgire a cadrului de investigare. Să presupunem că avem cele trei categorii de liceu şi ne interesează distribuirea lor, nu în legătură cu o singură facultate (cea de psihologie), ci în legătură cu trei tipuri de facultăți: „umaniste", „artistice" şi „tehnice".

Dacă realizăm un cadru de reprezentare sintetic al valorilor celor două variabile, obținem ceea ce se numeşte un tabel de corespondență. Iată cum ar arăta un astfel de tabel, pentru un set de date ipotetice:

	Liceu umanist	Liceu real	Liceu artistic	**Total pe linii**
Fac. Umaniste	45	20	30	**95**
Fac. Tehnice	14	60	12	**86**
Fac. Artistice	20	13	50	**83**
Total pe coloane	**79**	**93**	**92**	**264**

Acesta este un tabel de corespondență pentru două variabile nominale, fiecare având câte trei valori distincte (categorii)[4]. Valorile din celule reprezintă numărul de cazuri (*frecvențele observate*) care corespund fiecărei combinații dintre categoriile celor două variabile. „Totalul pe linii" exprimă numărul de studenți din fiecare facultate, consemnați în cercetare, indiferent de tipul de liceu absolvit, „totalul pe coloane", exprimă numărul de absolvenți din fiecare tip de liceu, indiferent de facultatea la care sunt înscrişi, iar la intersecția celor două totaluri regăsim totalul general al subiecților cercetării (N=264).

Fundamentarea testului statistic

Având un număr de 95 de studenți în „facultăți umaniste", această înseamnă că ei reprezintă 36% din totalul subiecților cercetării (95/264*100=36). Acest procent indică se referă la absolvenții care au ales o facultate de tip umanist, indiferent de liceul absolvit. În mod similar, calculăm procentele corespunzătoare celorlalte tipuri de facultăți. Valorile astfel calculate, pentru fiecare linie a tabelului, se numesc frecvențe marginale.

Dacă alegerea facultății nu ar avea nici o legătură cu tipul de liceu absolvit atunci, în mod normal, ar trebui să regăsim, pentru fiecare tip de liceu, acelaşi procent care exprimă ponderea studenților din fiecare facultate în totalul subiecților cercetați. Având procentele studenților din fiecare facultate şi numărul absolvenților din fiecare tip de liceu, putem calcula frecvențele „teoretice" (aşteptate) pentru fiecare celulă a tabelului. De exemplu, dintre cei 79 de absolvenți de liceu umanist consemnați de cercetare, 36% ar trebui să se afle în facultăți umaniste, ceea ce înseamnă: (79*36)/100=*28.4*. În mod similar, ar trebui să avem 32.5% (*25.6*) în facultăți ştiințifice şi 31.5% (*24.8*) în facultăți artistice. Acelaşi raționament se aplică mai departe şi celorlalte tipuri de liceu, cu utilizarea procentului corespunzător fiecărei facultăți. Precizăm că frecvențele teoretice (aşteptate) vor fi aceleaşi, în fiecare celulă, chiar dacă vor fi calculate pe baza frecvențelor marginale de pe coloane.

	Liceu umanist	Liceu real	Liceu artistic	**Total pe linie**	% pe linii
Fac. Umaniste	45 *(28.4)*	20 *(33.4)*	30 *(33.1)*	**95**	(95/264)*100=**36%**
Fac. Tehnice	14 *(25.6)*	60 *(30.2)*	12 *(29.9)*	**86**	(86/264)*100=**32,5%**
Fac. Artistice	20 *(24.8)*	13 *(29.2)*	50 *(28.9)*	**83**	(83/264)*100=**31.5%**
Total pe coloană	**79**	**93**	**92**	**264**	

[4] În mod similar, se pot crea tabele de corespondență pentru variabile categoriale având, fiecare, un număr diferit de valori (categorii).

Copyright © DEPARTAMENT ID 2008

Aşa cum constatăm, între frecvențele observate şi cele aşteptate sunt diferențe. Suma frecvențelor aşteptate (teoretice) este egală cu suma frecvențelor observate (poate rezulta o anumită diferență între totaluri, ca urmare a aproximării zecimalelor).

În final, problema cercetătorului este aceea de a stabili dacă între frecvențele observate şi cele teoretice (calculate) este o diferență care să justifice aprecierea că între cele două variabile există sau nu o legătură. Datele de acest gen nu mai pot fi analizate prin prisma distribuției binomiale, deoarece implică mai mult decât două „evenimente" posibile. De aceea, distribuția acestora se numeşte „distribuție multinomială". Desigur procedura de calcul pentru acest caz ar putea urma modelul celei binomiale dar, din cauza complexității ei, s-a apelat la o soluție mai simplă. Această soluție este fundamentată pe o aproximare derivată din formula binomială a lui *z*, care este pur si simplu ridicată la pătrat, devenind:

$$z^2 = \frac{(X - N * P)^2}{N * P * Q} \quad \text{(formula 4.8)}$$

Dacă înainte de ridicarea la pătrat ***z*** urmează o distribuție normală, după ridicarea la pătrat z urmează un alt tip de distribuție, numită „*chi-pătrat*", simbolizată cu litera grecească χ cu indicele de ridicare la pătrat (χ^2). Fără a intra în amănunte, vom preciza că distribuția χ^2 prezintă următoarele caracteristici:

- este, la fel ca şi distribuția normală, o familie de distribuții;
- are formă asimetrică;
- are originea în zero (din cauza ridicării la pătrat);
- are o formă dependentă de numărul de grade de libertate.

La fel ca şi distribuțiile ***t*** şi F, distribuția χ^2 este dependentă de numărul gradelor de libertate.

Acestea se calculează pe baza tabelului de corespondență dintre cele două variabile, astfel:

df=(număr coloane-1)*(număr linii-1)

Formula de calcul pentru testul *chi-pătrat*, derivată din formula 4.8, este :

$$\chi^2 = \sum \frac{(f_O - f_E)^2}{f_E} \quad \text{(formula 4.9)}$$

unde f_O este frecvența observată, iar f_E, frecvența aşteptată.

Decizia pentru testul *chi-pătrat* se bazează pe compararea valorii calculate cu o valoare critică, corespunzătoare nivelului alfa ales (0.05 sau, opțional, mai mic). Valorile critice pentru distribuția *chi-pătrat* se găsesc într-o tabelă specială (vezi anexa 6). Dacă valoarea calculată a lui χ^2 este egală sau mai mare decât valoarea critică pentru nivelul ales al lui alfa, atunci ipoteza de nul poate fi respinsă, iar ipoteza cercetării confirmată.

Pe această structură formală se bazează două variante distincte ale testului *chi-pătrat*: testul corespondenței (*Goodness of Fit*) şi testul asocierii. Primul, compară frecvențele observate ale valorilor unei singure variabile cu frecvențele aşteptate pentru acele valori. Al doilea, compară frecvențele valorilor observate pentru două variabile cu frecvențele lor aşteptate, cu scopul de a testa relația (asocierea) dintre cele două variabile.

Testul chi-pătrat pentru gradul de corespondență (*goodness of fit*)

Această variantă a testului *chi-pătrat* compară frecvențele observate ale unei distribuții cu frecvențele teoretice (aşteptate) ale acelei variabile. De exemplu, dacă avem frecvențele unei variabile putem afla dacă aceasta se distribuie după curba normală (z), prin compararea cu frecvențele cunoscute ale acestei distribuții (aria de sub curbă).

Copyright © DEPARTAMENT ID 2008

Să presupunem că a fost aplicat un test de cunoştinţe unui eşantion de 200 de elevi, care a fost evaluat cu calificative, astfel: *F.Slab, Slab, Mediu, Bun, F.Bun*.

Problema cercetării: Calificativele obţinute se distribuie normal la nivelul clasei?

Populaţia 1: Calificativele obţinute de elevi.

Populaţia 2: Calificativele, aşa cum s-ar distribui pe o curbă normală: FS=2.5%, B=14%, M=67%, B=14% şi FB=2.5% (procentele sunt cele tipice unei curbe *z*, împărţite în cinci clase valorice).

- *Ipoteza cercetării (H_1)*: Distribuţia calificativelor urmează legea curbei normale la nivelul eşantionului de elevi.
- *Ipoteza de nul (H_0)*: Distribuţia calificativelor nu urmează legea curbei normale în rândul elevilor examinaţi.

Determinarea caracteristicilor deciziei statistice:

- alegem α=0.05 (în cazul testului χ^2 decizia nu poate fi decât unilaterală, deoarece acest test nu poate lua valori negative)
- găsim valoarea critică pentru χ^2=9.48 în tabela pentru distribuţia χ^2, pentru df=(2-1)*(5-1)=4 şi α=0.05

Tabelul următor conţine datele de cercetare şi algoritmul de calcul:

Calificativ	**Frecvenţa observată (f_O)**	**Frecvenţa aşteptată (f_E)**	$\frac{(f_O - f_E)^2}{f_E}$
FB	10	2.5% din 200 =5	$\frac{(10-5)^2}{5} = 5.00$
B	34	14% din 200 =28	$\frac{(34-28)^2}{28} = 1.29$
M	140	67% din 200 =134	$\frac{(140-134)^2}{134} = 0.27$
S	10	14% din 200 =28	$\frac{(10-28)^2}{28} = 11.57$
FS	6	2.5% of 200 =5	$\frac{(6-5)^2}{5} = 0.20$
Σ	200	-	$\chi^2 = \sum \frac{(f_O - f_E)^2}{f_E} = 18.33$

Decizia statistică:

- χ^2 calculat (18,33) este mai mare decât χ^2 critic (9,48)
- Respingem ipoteza de nul şi tragem concluzia că distribuţia calificativelor urmează forma curbei normale.

Concluzia statistică poate fi interpretată, în acest caz, ca fiind negativă din punctul de vedere al eficienţei procesului didactic. În mod normal, dacă activitatea de învăţare ar fi eficientă, rezultatele elevilor ar trebui să se distribuie asimetric negativ, adică cu tendinţă de grupare a valorilor spre calificativele superioare. Rezultatele procesului de învăţare nu se distribuie „normal", nefiind un proces „natural", ci unul în care valorile (calificativele) sunt supuse unei influenţe sistematice (prin efortul profesorilor şi al elevilor înşişi) înspre valorile mari.

Facem, încă o dată, precizarea că această formă a testului *chi-pătrat* se aplică atunci când vrem să comparăm frecvenţe observate cu frecvenţe teoretice (aşteptate), pe care le cunoaştem deja. El este echivalentul testului *z* pentru proporţii pentru distribuţia binomială, cu specificaţia că se utilizează atunci când avem mult de două categorii. Testul *chi-pătrat* pentru gradul de corespondenţă *(goodness of fit)* nu are un indice de mărime a efectului.

Copyright © DEPARTAMENT ID 2008

Iată câteva exemple posibile de cercetări ale căror date pot fi analizate cu testul *chi-pătrat* al gradului de corespondență:

- Vrem să ştim dacă există o preferință pentru o anumită categorie de muzică (clasică, populară, pop-rock). În acest caz, dacă distribuția preferințelor nu ar fi influențată de nici o anumită preferință (ipoteza de nul) atunci frecvența aşteptată (teoretică) pentru fiecare gen muzical ar trebui să fie echivalentă cu 100/3=33.3% numărul subiecților. Mai departe, nu ne rămâne decât să testăm diferența dintre cele două categorii de frecvențe (teoretice şi observate), conform modelului de calcul de mai sus.
- Într-un studiu asupra relației dintre atractivitate şi preferința pentru profesori, unui număr de studenți li se prezintă fotografiile preselectate ale unor şase potențiali profesori, ale căror portrete sugerează grade diferite de atractivitate, şi li se cere să aleagă dintre aceştia pe cel pe care ar dori să îl aibă ca profesor. Dacă gradul de atractivitate nu are nici un impact asupra preferinței ca profesor, atunci frecvențele cu care sunt alese fotografiile ar trebui să fie egale (100/6=16.6%).
- Într-un studiu de marketing, o companie trebuie să aleagă dintre patru propuneri imagini. Acestea sunt prezentate unui eşantion de subiecți şi se consemnează numărul de preferințe exprimate pentru fiecare imagine. Dacă toate ar avea acelaşi impact, atunci numărul de preferințe ar trebui să fie egal (25%, pentru fiecare imagine).

Sarcina de lucru nr. 3. 4

Vrem să ştim dacă există o preferință pentru un anumit gen muzical (clasică, populară, dance, rock) printre cei 200 de studenți ai facultății de psihologie. În acest caz, dacă distribuția preferințelor nu ar fi afectată de nici o preferință (ipoteza de nul), atunci frecvența aşteptată (teoretică) pentru fiecare din cele patru genuri muzicale este 25% numărul subiecților (100% împărțit la cele patru genuri muzicale), adică exact 50 pentru fiecare gen muzical. În realitate, preferințele exprimate de studenți se distribuie astfel: muzică clasică=30; muzică populară=15; muzică dance=100; muzică rock=55. Problema cercetării este dacă există o distribuție preferințială pentru genurile muzicale, în rândul studenților de la psihologie (pentru alfa=0.5).[5]

Verifică răspunsul corect la pagina 22

Chi-pătrat - testul asocierii (independence chi-square)[6]

Această variantă a testului *chi-pătrat* este mai frecvent utilizată. Ea compară frecvențele observate ale unei distribuții (variabile) cu frecvențele corespondente ale altei distribuții (variabile), ambele măsurat pe scale de tip categorial, cu scopul de a vedea dacă există o relație între cele două variabile. Să presupunem că avem rezultatele la testul de statistică (măsurate pe o scală ordinală şi notate, convențional, cu A, B, C, D, E, unde A reprezintă nivelul de performanță cel mai ridicat iar E, cel mai scăzut).

Problema cercetării: Dorim să aflăm dacă există o diferență semnificativă între băieți (M) şi fete (F) la testul de statistică.

Ipoteza cercetării: Distribuția performanței depinde de genul „masculin" sau „feminin".

Ipoteza de nul: Rezultatele la testul de statistică nu au legătură cu variabila sex.

Determinarea criteriilor de decizie statistică:

- alegem α=0.05
- df=(2-1)*(5-1)=4
- citim valoarea critică pentru χ^2 în tabela pentru distribuția χ^2:

[5] Datele sunt fictive.

[6] Cunoscut şi sub numele „*testul chi-pătrat Pearson al asocierii*", a fost elaborat de Karl Pearson.

Copyright © DEPARTAMENT ID 2008

- χ^2_{critic}= 9.49

Datele cercetării ar putea fi astfel centralizate în următorul tabel de corespondență[7]:

	Performanța la test					
	A	**B**	**C**	**D**	**F**	**Total**
Masculin	10	34	140	10	6	**200** = 57.14% din total general
Feminin	10	32	97	6	5	**150** = 42.86% din total general
Total	**20**	**66**	**237**	**16**	**11**	**Total general=350**

- Frecvențele marginale sunt: 200 (57.14%) pentru „băieți" şi 150 (42.86%) pentru „fete"
- Dacă performanța la test nu are nici o legătură cu genul subiecților, trebuie să regăsim aceste procente pentru fiecare dintre calificativele acordate.
- Aceasta înseamnă că, teoretic, în celula **A/Masculin,** ar trebui să găsim, proporțional, tot atâția băieți câți sunt pe întregul lot (57.14%). Adică (20*57.14)/100=11.42, care reprezintă frecvența aşteptată pentru celula respectivă din tabelul de corespondență.
- La fel, pentru celula **A/Feminin** ar trebui să avem 42.86% din totalul pentru „feminin", adică: (20*42.86)/100=8.52.
- În acelaşi mod de calculează frecvențele observate pentru fiecare celulă a tabelului.

Pentru o mai uşoară înțelegere a mecanismului de calcul, vom rearanja tabelul astfel:

Celule	**Frecvența observată (f_O)**	**Frecvența aşteptată**	$\frac{(f_O - f_E)^2}{f_E}$
Masculin – A	10	(20*57.14)/100=11.43	$\frac{(10-11.43)^2}{11.43} = 0.18$
Masculin – B	34	(66*57.14)/100=37.71	$\frac{(34-37.71)^2}{37.71} = 0.36$
Masculin – C	140	(237*57.14)/100=135.42	$\frac{(140-135.42)^2}{135.42} = 0.15$
Masculin – D	10	(16*57.14)/100=9.14	$\frac{(10-9.14)^2}{9.14} = 0.08$
Masculin – F	6	(11*57.14)/100=6.29	$\frac{(6-6.29)^2}{6.29} = 0.01$
Feminin – A	10	(20*42.86)/100=8.57	$\frac{(10-8.57)^2}{8.57} = 0.24$
Feminin – B	32	(66*42.86)/100=28.29	$\frac{(32-28.29)^2}{28.29} = 0.49$
Feminin – C	97	(237*42.86)/100=101.58	$\frac{(97-101.58)^2}{101.58} = 0.21$
Feminin – D	6	(16*42.86)/100=6.86	$\frac{(6-6.86)^2}{6.86} = 0.11$
Feminin – F	5	(11*42.86)/100=4.71	$\frac{(5-4.71)^2}{4.71} = 0.02$
Σ	**350**		$X^2 = \sum \frac{(f_O - f_E)^2}{f_E} = 1.85$

[7] Datele din acest exemplu nu se referă la o situație reală.

Copyright © DEPARTAMENT ID 2008

- Se compară χ^2 critic (9.49) cu χ^2 calculat (1.85) pentru df = (2-1)(5-1) = 4
- Valoarea calculată a testului este mai mică decât valoarea critică, ca urmare, acceptăm ipoteza de nul. Rezultatele la test nu confirmă ipoteza că rezultatele se distribuie în funcție de apartenența de gen a subiecților.

Condiții pentru aplicarea testului χ^2

- Cele două variabile nu trebuie să se „intersecteze" (să nu existe subiecți care să fie incluşi în mai mult de o celulă de tabel)
- Selecție aleatoare a eşantioanelor
- Este recomandabil ca frecvența aşteptată să nu ia valori mai mici de 5 (sau, cel puțin, în nu mai mult de 20% din celule).
- Nici o celulă nu trebuie să aibă frecvența aşteptată mai mică de 1.

Pentru situațiile în care frecvențele aşteptate sunt mai mici decât specificațiile de mai sus, sau atunci când tabelul de corespondență dintre variabile are două linii şi două coloane, se recomandă aplicarea unei corecții la formula de bază. Aceasta se numeşte „corecția Yeates" şi constă în scăderea unei constante (0.5) din expresia de la numărător, luată în valoare absolută:

$$X^2 = \sum \frac{\left(|f_O - f_E| - 0.5\right)^2}{f_E} \quad \text{(formula 4.10)}$$

Utilizarea testului *chi-pătrat* al asocierii

Testul *chi-pătrat* al asocierii se utilizează atunci când dorim să testăm relația dintre două variabile, ambele măsurate pe scală de tip categorial. Facem precizarea că variabilele categoriale deşi sunt, de regulă, de tip nominal, pot fi atât ordinale cât şi de interval sau de raport. Ceea ce caracterizează o variabilă categorială nu este atât scala de măsurare, cât faptul că primeşte puține valori, care împart distribuția în categorii de valori. De exemplu, într-un studiu cu privire la relația dintre gravitatea accidentelor de circulație („fără răniți", „cu răniți uşor", „cu răniți grav", „cu morți") şi puterea motoarelor (1400 cm^3, 1600 cm^3, 2000 cm^3, 2500 cm^3, 3000 cm^3), ambele variabile sunt de tip categorial, dar prima este pe scală nominală, iar a doua pe scală cantitativă.

Testul *chi-pătrat* al asocierii (independenței) poate fi văzut ca un veritabil test de corelație pentru date categoriale. De asemenea, poate fi folosit în locul testului *t* sau ANOVA, dacă nu sunt îndeplinite condițiile pentru variabila dependentă. Într-un asemenea caz, variabila dependentă cantitativă se transformă, prin gruparea în frecvențe, în variabilă de tip categorial. Această opțiune se va alege numai dacă ne aflăm în fața unei flagrante violări a condiției de normalitate, deoarece testele parametrice au o putere mai mică decât cele neparametrice. La fel ca şi în cazul altor teste statistice, nu se vor putea trage concluzii de tip cauzal decât numai dacă variabilele sunt măsurate în contextul unui experiment psihologic.

Marimea efectului pentru testul *chi pătrat* al asocierii

Coeficientul *φ (fi)*

Atunci când utilizăm testul pentru asocierea variabilelor, valoarea χ^2 certifică faptul că cele două variabile sunt relaționate. Dar mărimea lui χ^2 nu ne spune nimic cu privire la intensitatea relației dintre variabile. De fapt, mărimea lui χ^2 este în funcție de N. Dacă multiplicăm frecvențele celulelor cu o constantă, valoarea lui χ^2 se multiplică şi ea cu acea constantă, singura consecință fiind aceea că se diminuează probabilitatea ca valoarea respectivă să fie obținută din întâmplare. Pentru completarea interpretării valorii χ^2 este necesar un indicator suplimentar, care să ne spună ceva şi despre intensitatea legăturii, nu doar despre semnificația acesteia. Un astfel de indicator este coeficientul φ (fi), care se calculează pentru asocierea variabilelor care prezintă fiecare doar două valori posibile (tabele de contingență 2x2).

Formula după care se calculează este:

Copyright © DEPARTAMENT ID 2008

$$\varphi = \sqrt{\frac{\chi^2}{N}} \qquad \text{(formula 4.11)}$$

Coeficientul *φ Cramer*

Coeficientul φ este adecvat doar pentru tabelele de contingenţă de tip 2x2, când ambele variabile sunt dihotomice. O uşoară modificare a acestuia, denumită *φ Cramer,* îl face utilizabil pentru intensitatea asocierii dintre variabile având un număr diferit de categorii.

Indicele *φ Cramer* se calculează după formula:

$$\varphi_c = \sqrt{\frac{\chi^2}{N*(L-1)}} \qquad \text{(formula 4.11 bis)}$$

unde:

- N este volumul eşantionului
- L este valoarea cea mai mică dintre numărul liniilor sau al coloanelor tabelului de corespondenţă (de exemplu, pentru un tabel de corespondenţă 4x3 - patru linii şi patru coloane - L are valoarea 3-1=2).

În cazul coeficienţilor φ, dacă frecvenţele fiecărei celule din tabelul de corespondenţă sunt multiplicate cu o constantă, atât χ^2 cât şi N cresc concomitent, iar valoarea coeficientului φ rămâne aceeaşi. Coeficientul φ se modifică numai dacă se modifică şi raporturile dintre proporţii, ceea ce înseamnă că mărimea lui nu este influenţată de N. El reprezintă un indicator numeric al intensităţii relaţiei şi poate lua valori între zero - absenţa relaţiei şi unu - relaţie perfectă între cele două variabile. De exemplu, pentru testul *chi-pătrat* al asocierii dintre gen şi performanţa la testul de statistică (care a rezultat nesemnificativ), al cărui tabel de corespondenţă este de forma 2x5, valoarea coeficientului φ_c este:

$$\varphi_c = \sqrt{\frac{\chi^2}{N*(2-1)}} = \sqrt{\frac{1.85}{350}} = 0.07$$

Interpretarea coeficienţilor φ

Valoarea coeficientului φ se asociază interpretării testului *chi-pătrat*, atunci când acesta este semnificativ, pentru a adăuga o informaţie suplimentară cu privire la intensitatea relaţiei. Prin ridicarea la pătrat a expresiei de calcul, coeficientul φ^2 poate fi interpretat procentual, la fel ca şi coeficientul de determinare (r^2), indicând proporţia variaţiei unei variabile determinată de variaţia celeilalte variabile. În cazul nostru, numai 0.4% (0.07^2*100) din variaţia calificativelor la testul de statistică este explicată prin diferenţa de gen (masculin/feminin), ceea ce, în conformitate cu decizia statistică, s-a dovedit a fi nesemnificativ.

În conformitate cu recomandările lui Cohen, cit. de Kotrlik şi Williams (2003), valorile lui φ vor fi interpretate după cum urmează:

φ (Cohen)	0.10	efect mic
	0.25	efect mediu
	0.40	efect mare

Raportarea rezultatului

În cazul testului χ^2 elementele care vor fi incluse în raport sunt următoarele: gradele de libertate, valoare testului, nivelul p şi coeficientul *φ* sau *Cramer φ*. În varianta narativă, pentru exemplul de mai sus, prezentarea rezultatelor ar putea avea următoarea formă:

„*Rezultatele testului de statistică, evaluate pe cinci clase valorice (A,B,C,D,E) au fost comparate pe sexe. Testul χ^2 pentru asocierea variabilelor indică faptul că rezultatele nu diferă semnificativ în funcţie de gen, $\chi^2(4) = 1.85$, p >0 .05, cu un coeficient φ=0.07, care indică o asociere slabă*”.

Copyright © DEPARTAMENT ID 2008

În cazul în care testul ar fi fost semnificativ, raportarea rezultatelor ar fi trebuit să conțină şi referințe cu privire la procentele consemnate în celulele tabelului de corespondență, astfel încât să fie scoase în evidență diferențele releavnte dintre categoriile comparate.

Sarcină de lucru nr. 3. 5

80 de pacienți depresivi au fost împărțiți aleatori în patru grupuri egale, fiecare dintre acestea urmând un tip diferit de psihoterapie (psihodinamică, rogeriană, de grup şi comportamentală). După şase luni de terapie, fiecare pacient a fost clasificat într-una din următoarele trei categorii: ameliorat, agravat, neschimbat. Datele sunt cele din tabelul următor:

Frecvențe observate	**model psihoterapeutic**				**suma pe linie**
	psihodinamic	**rogerian**	**de grup**	**comportamental**	
ameliorat	6	4	8	12	**30**
neschimbat	6	14	3	5	**28**
înrăutățit	8	2	9	3	**22**
suma pe coloane	**20**	**20**	**20**	**20**	**N=80**

Problema cercetării: Pe baza datelor existente, se poate concluziona că există o diferență de eficiență le celor patru tehnici terapeutice asupra depresiei? (pentru alfa=0.5).

Verificați răspunsurile corecte

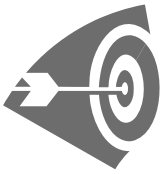
Testul exact Fisher

Aşa cum am precizat, testul *chi-pătrat* este calculat pe baza unei formule ale cărei rezultate nu urmează cu maximă precizie distribuția χ^2. Dacă în cele mai multe situații acest lucru nu reprezintă un neajuns notabil, sunt si cazuri în care rezultatele pot fi alterate suficient de mult pentru a putea fi luate în considerare:

- atunci când volumul eşantionului este redus (N<20);
- atunci când valorile f_e pentru una sau mai multe dintre celulele tabelei de corespondență sunt foarte mici.

În aceste situații, precum şi atunci când tabelul de corespondență este compus din două linii şi două coloane, este recomandabilă utilizarea testului exact Fisher. El se bazează pe calcularea tuturor tabelelor posibile ce pot fi construite pentru frecvențele marginale. Deoarece necesită un mare volum de calcule, testul exact Fisher se efectuează numai cu ajutorul programelor computerizate.

Copyright © DEPARTAMENT ID 2008



Rezumatul unității de învățare

- Distribuția binomială derivă din serii de evenimente independente dihotomice. Cele două posibilități ale fiecărui eveniment au probabilitățile P şi Q, a căror sumă este 1 (de unde Q=1-P).
- Atunci când P=Q=0.5, distribuția binomială este simetrică. Pe măsură ce numărul evenimentelor (N) creşte, distribuția binomială se apropie de forma normală. Chiar şi atunci când P≠Q distribuția binomială se apropie de forma normală odată cu creşterea lui N.
- Atunci când N creşte la infinit, distribuția binomială devine normală, având media=N*P şi abaterea standard= $\sqrt{N*P*Q}$. Ca urmare, probabilitatea ca un anume eveniment să cadă în categoria P poate fi aproximată prin calcularea unui scor z şi evaluarea ariei corespunzătoare de sub curba normală.
- Dacă P=0.5, distribuția normală devine o aproximare bună pentru distribuția normală începând cu N=25.
- Testul semnului poate fi utilizat în locul testului t pentru eşantioane dependente atunci când nivelul diferenței dintre cele două determinări nu poate fi evaluat, ci numai direcția diferenței. Dat fiind faptul că fiecare diferență poate fi într-una din categorii (+ sau -) distribuția binomială poate fi utilizată pentru a estima în ce măsură dezechilibrul între cele două categorii este posibil să apară din întâmplare (prin raportare la distribuția normală).
- Atunci când N nu este foarte mare, utilizarea distribuției normale pentru aproximarea distribuției binomiale introduce o eroare sistematică care poate fi compensată prin corecția de continuitate, extrăgând 0.5 din valoare absolută a diferenței de la numărătorul scorului z.
- Dacă evenimentele probabilistice pot avea mai mult decât două posibilități (de ex., adevărat-fals), probabilitatea cu care fiecare eveniment cade într-una din categoriile posibile se supune distribuției multinomiale.
- Din cauza complexității procesului de evaluare a probabilităților multinomiale, este utilizată o estimare a acestora prin distribuția chi-pătrat. Numărul gradelor de libertate pentru distribuția multinomială este dat de numărul categoriilor minus 1.
- Testul chi-pătrat are două variante: (1) Testul chi-pătrat al asocierii testează diferența dintre valorile a două variabile categoriale (nominale sau ordinale). (2) Testul chi pătrat al corespondenței (*goodness-of-fit*) măsoară diferența (“potrivirea”) dintre valorile unei variable categoriale şi probabilitățile teoretice dinainte cunoscute ale acestor valori.
- Diferențele mari dintre frecvențele observate şi cele aşteptate produc valori ridicate ale testului chi-pătrat, care cad în zona dreaptă (pozitivă) a distribuției de nul şi conduc la respingere a ipotezei de nul. Diferențele mici, produc valori ale testulu chi-pătrat apropiate de zero, conducând la acceptarea ipotezei de nul.
- Atunci când fiecare dintre cele două variabile au doar două categorii, situație în care frecvențele aşteptate sunt prea mici pentru a justifica o estimare chi-pătrat, se utilizează testul exact Fischer.

Copyright © DEPARTAMENT ID 2008

Răspunsuri şi comentarii la problemele de evaluare

Răspunsuri corecte la sarcină de lucru nr. 3. 1

(1) Se acceptă ipoteza de nul (pentru alfa=0.05, bilateral). Studentul a răspuns la întâmplare.
$z = \frac{58 - 50}{\sqrt{100 * 0.5 * 0.5}} = 1.6$ (mai mic decât 1.96, cât este z critic pentru alfa=0.05 bilateral)

(2) (a) $z = \frac{0.22 - 0.15}{\sqrt{\frac{0.85 * 0.15}{120}}} = \frac{0.07}{0.03} = 2.33$

(b) da, deoarece z calculat este mai mare decât z critic pentru alfa 0.05, bilateral (1.96)

Răspuns la sarcina de lucru 3.2

$$z = \frac{\left(0{,}25 - \frac{1}{2 * 30}\right) - \left(0.60 - \frac{1}{2 * 30}\right)}{\sqrt{\frac{0.25 * 0.75}{30} + \frac{0.60 * 0.40}{30}}} = \frac{0.24 - 0.59}{\sqrt{0.006 + 0.008}} = \frac{-0.35}{0.11} = -3.18$$

Se respinge ipoteza de nul. Stresul temporal reduce performanța în rezolvarea de probleme.

Răspuns la sarcina de lucru 3.3

(1) $z = \frac{|30 - 35 * 0.5| - 0.5}{\sqrt{35 * 0.5 * 0.5}} = \frac{12}{2.95} = 4.06$

z calculat (4.06) este mai mare decât z critic pentru alfa=0.05 bilateral (1.96). Ca urmare, se respinge ipoteza de nul şi se acceptă ipoteza cercetării. Discuțiile în grup au efect asupra opiniei individuale.

(2) p=(0.0938+0.0156)*2=0.2188 Se acceptă ipoteza de nul (se respinge ipoteza cercetării), deoarece p=0.21 este mai mare decât nivelul alfa=0.05.

Răspuns la sarcina de lucru 3.4

Gen muzical	**Frecvența observată (f_O)**	**Frecvența aşteptată (f_E)**	$\frac{(f_O - f_E)^2}{f_E}$
clasică	30	25% din 200 =50	$\frac{(30-50)^2}{50} = 8$
populară	15	25% din 200 =50	$\frac{(15-50)^2}{50} = 24.5$

Copyright © DEPARTAMENT ID 2008

	dance	100	25% din 200 =50	$\frac{(100-50)^2}{50}=50$
	rock	55	25% din 200 =50	$\frac{(55-50)^2}{50}=0.5$
	Σ	200	-	83

$$\chi^2=\Sigma\frac{(f_O-f_E)^2}{f_E}=83$$

Criterii de decizie: alfa=0.05; df=(2-1)*(5-1)=4
Valoarea critică pentru $X^2=9.49$ (atenție X^2 nu se testează bilateral)

Decizia: X2 calculat (83) este mai mare decât X2 critic (9.49), iar ca urmare, se respinge ipoteza de nul. Studenții de la psihologie au anumite preferințe în raport cu cele patru genuri muzicale.

Rezultatul sarcinii de lucru 1.5

Criterii de decizie: df=(3-1)*(4-1)=6; alfa=0.05; X^2 critic=12.59

Pentru uşurința calculelor, datele se aranjează ca în tabelul următor:

	fo	f_E	$\frac{(f_O-f_E)^2}{f_E}$
ameliorat-I	6	(30/80)*20=7.5	$(6-7.5)^2/7.5=0.3$
ameliorat-II	4	(30/80)*20=7.5	$(4-7.5)^2/7.5=1.63$
ameliorat-III	8	(30/80)*20=7.5	$(8-7.5)^2/7.5=0.03$
ameliorat-IV	12	(30/80)*20=7.5	$(12-7.5)^2/7.5=2.7$
neschimbat-I	6	(28/80)*20=7	$(6-7)^2/7=0.14$
neschimbat-II	14	(28/80)*20=7	$(14-7)^2/7=7$
neschimbat-III	3	(28/80)*20=7	$(3-7)^2/7=2.28$
neschimbat-IV	5	(28/80)*20=7	$(5-7)^2/7=0.57$
înrăutățit-I	8	(22/80)*20=5.5	$(8-5.5)^2/5.5)=1.13$
înrăutățit-II	2	(22/80)*20=5.5	$(2-5.5)^2/5.5)=2.22$
înrăutățit-III	9	(22/80)*20=5.5	$(9-5.5)^2/5.5)=2.22$
înrăutățit-IV	3	(22/80)*20=5.5	$(3-5.5)^2/5.5)=1.13$
			Σ=21.4

Valoarea calculată ($X^2=21.4$) este mai mare decât valoarea X^2 critică (9.49). Ca urmare, ipoteza de nul se respinge.

Concluzia cercetării: Există o diferență între cele patru categorii psihoterapeutice în ceea ce priveşte eficiența asupra depresiei.[8]

[8] Datele cercetării sunt fictive.

Copyright © DEPARTAMENT ID 2008

Lucrarea de evaluare nr. 3.1 şi modul de cotare

Lucrarea de evaluare va vi publicată pe portal. Termenul de trimitere este preziua tutorialului. După această dată lucrările trimise nu vor mai fi acceptate.

Copyright © DEPARTAMENT ID 2008

TESTE STATISTICE NEPARAMETRICE PENTRU DATE ORDINALE

Obiectivele unității de învățare şi informații introductive

Parcurgerea acestei unități va permite studenților:

- ✓ Să calculeze testul Mann-Whitney
- ✓ Să calculeze testul Kruskall-Wallis
- ✓ Să calculeze testul Wilcoxon
- ✓ Să calculeze testul Friedman
- ✓ Să calculeze testul de corelație pentru date ordinale Spearman

Introducere

Testele statistice pentru date ordinale se utilizează în două situații:

a) Atunci când variabila dependentă este exprimată pe scală de tip ordinal. În acest caz valorile nu au proprietăți de interval dar exprimă poziția fiecăreia în raport cu cealaltă.

b) Atunci când variabila dependentă este măsurată pe scală de interval/raport dar distribuția ei nu respectă condițiile impuse de testele parametrice. În această situație se efectuează transformare de rang, adică înlocuieşte fiecare valoare a distribuției cu poziția pe care o are în cadrul distribuției, sub aspectul ordinii de mărime. Noua distribuție rezultată poate fi supusă analizei statistice cu teste neparametrice ordinale.

Testul Mann-Whitney (U) pentru două eşantioane independente

Să luăm în considerare următoarea problemă: Un psiholog care lucrează într-o mare bancă doreşte să vadă dacă există o diferență între premiile băneşti anuale primite de femeile şi bărbații angajați ai băncii. În tabelele de mai jos se află nivelurile primelor şi rangurile acestora în raport cu întreaga distribuție a primelor, indiferent de sex.

Copyright © DEPARTAMENT ID 2008

Masculin	„Premiu” (mil. lei)	Rang „Primă”
1	9	26
1	34	22
1	35	21
1	43	18.5
1	56	14
1	61	12
1	62	11
1	64	10
1	67	9
1	67	9
1	70	7
1	75	6
1	80	5
1	87	4
1	88	3
1	110	2
1	200	1
n_A=17		ΣR_A=180.5

Feminin	„Premiu” (mil. lei)	Rang „primă”
2	3	27
2	17	25
2	26	24
2	32	23
2	36	20
2	43	18.5
2	44	17
2	47	16
2	51	15
2	59	13
n_B=10		ΣR_B=198.5

Problema este una tipică pentru a fi rezolvată cu testul ***t*** al diferenței dintre mediile a două eşantioane independente. Avem o variabilă independentă de tip nominal-dihotomic şi una dependentă, de tip interval/raport. Din păcate, analiza preliminară a variabilei dependente („primă”) relevă abateri mari de la condițiile de normalitate (un indice de boltire de peste 7) precum şi o slabă reprezentativitate a mediei, ambele datorate, mai ales, prezenței unei valori extreme (o primă de 200 mil. lei). După ce verificăm corectitudinea valorii respective, ajungem la concluzia că ea nu poate fi eliminată şi, ca urmare, nu este recomandabilă utilizarea unui test parametric.

Într-o situație de acest gen este aplicabil testul „Mann-Whitney U”[9] pentru date ordinale. Pe ultima coloană a fiecărui tabel găsim transformarea în ranguri a valorilor variabilei dependente. Atribuirea rangurilor în mod descrescător sau crescător este nerelevantă. Dacă toate valorile sunt distincte, fiecare valoare primeşte un rang distinct. Atunci când există valori identice, valorile respective primesc un rang egal cu media aritmetică a rangurilor cuvenite. Se poate alege şi soluția atribuirii tuturor valorilor identice primul rang cuvenit (ranguri *ex aequo*).

Procedura de calcul:

Se calculează cele două valori U, corespunzătoare grupurilor A (masculin) şi B (feminin), astfel:

$$U_A = n_A * n_B + \frac{n_A * (n_A + 1)}{2} - \Sigma R_A \qquad \text{(formula 5.1)}$$

respectiv,

$$U_B = n_A * n_B + \frac{n_B * (n_B + 1)}{2} - \Sigma R_B \qquad \text{(formula 5.2)}$$

[9] Desemnat uneori şi sub numele „Wilcoxon-Mann-Withney”, sau „testul U”.

Copyright © DEPARTAMENT ID 2008

Unde:
n_A şi n_B reprezintă volumul celor două grupuri independente care compun eşantionul
ΣR_A şi ΣR_B reprezintă suma rangurilor pentru fiecare din cele două grupuri

Pentru exemplu nostru:

$$U_A = 17*10 + \frac{17*18}{2} - 180.5 = 170 + 153 - 180.5 = 142.5$$

respectiv,

$$U_B = 17*10 + \frac{10*11}{2} - 198.5 = 170 + 55 - 198.5 = 26.5$$

Valoarea testului Mann-Whitney este dată de valoarea U cea mai mică, în cazul nostru U_B (26.5).

Decizia statistică se ia prin compararea valorii U celei mai mici cu valoarea citită în tabelul valorilor critice pentru testul Mann-Whitney U, în funcție de nivelul alfa, n_A şi n_B (Anexa 7).

În cazul testului U decizia statistică se ia astfel:

- Se respinge ipoteza de nul dacă valoarea U calculată este mai mică sau egală cu valoarea critică tabelară.
- Se acceptă ipoteza de nul dacă valoarea U calculată este mai mare decât valoarea critică tabelară.

În general tabelele de decizie pentru testul Mann-Whitney nu acoperă decât parțial situațiile posibile şi nu trec de valori ale lui n_A şi n_B mai mari de 20. Pentru exemplul nostru, valoarea critică corespunzătoare pentru $U_{0.05;17:10}$=48 (dacă preferăm aproximarea, mai conservatoare, n_A=18).

Deoarece $U_B < U_{0.05;17:10}$ ipoteza de nul se respinge şi se acceptă ipoteza cercetării. Nivelul primelor anuale este semnificativ diferit pentru bărbați față de femei[10].

Afirmam mai sus că tabelele statistice pentru testul Mann-Whitney U nu se referă la grupuri mai mari de 20. Aceasta deoarece, de la acest volum în sus, distribuția valorilor testului poate fi aproximată de curba normală z, iar testul poate fi calculat cu formula următoare:

$$z = \frac{\Sigma R_A - 0.5 * n_A * (N+1)}{\sqrt{n_A * n_B * (N+1)/12}} \quad \text{(formula 5.3)}$$

(unde $N=n_A+n_B$)

Valoarea lui ***z*** astfel obținută este comparată cu valorile critice tabelare de pe curba normală corespunzătoare nivelului alfa ales, unilateral sau bilateral.

Publicarea rezultatului

La publicarea rezultatului pentru testul Mann-Whitney U se vor indica:

- volumul grupurilor comparate (nA şi nB)
- valoarea testului (U)
- pragul de semnificație (*p*).

[10] Desigur, nu se poate invoca neapărat o discriminare de sex în acordarea primelor, dacă pozițiile profesionale ocupate de subiecții cercetării sunt diferite. Rezultatul poate sugera, însă, că bărbații ocupă poziții profesionale mai înalte decât femeile.

Copyright © DEPARTAMENT ID 2008

Sarcina de lucru nr. 3.6

Două grupuri de câte 5 subiecți au fost supuse unui experiment care vizează reacția fiziologică la prezentarea unui stimul şocant. Unul din grupuri a primit un instructaj care sugera natura şocantă a stimulului în timp ce celuilalt grup i s-a spus că vor primi stimuli agreabili. Datele de mai jos prezintă valoarea pulsului măsurată în primele 10 secunde de la prezentarea stimulului:

Grupul cu instructaj liniştitor:
105
130
145
125
115
Grupul cu instructaj pregătitor corect
120
80
90
110
95

Verificați răspunsurile corecte

Testul Kruskall-Wallis pentru mai mult de două eşantioane independente

Pentru evaluarea diferenței la nivel de ranguri între mai mult de două eşantioane independente se utilizează testul Kruskal-Wallis. Acesta poate fi asimilat unei analize de varianță pentru date ordinale.

Să presupunem că avem trei categorii de subiecți (piloți, controlori de trafic şi navigatori de bord) cărora le-a fost aplicat un test de reprezentare spațială. Să presupunem, de asemenea, că valorile variabilei dependente nu se pretează la un test ANOVA, dată fiind prezența câtorva valori extreme ce nu pot fi eliminate. În aceste condiții, testul Kruskal-Wallis este alegerea potrivită. Aceasta presupune ordonarea după rang a valorilor variabilei dependente (reprezentare spațială) pentru toate categoriile de subiecți, luate împreună.

Copyright © DEPARTAMENT ID 2008

Grup	rang
1	2
1	6
1	7
1	11
1	12
1	3
2	5
2	8
2	10
3	1
3	4
3	9
3	13

Tabelul prezintă datele cercetării. Variabila „grup" este una de tip nominal, fiecare din cele trei grupuri fiind codificat cu o valoare convențională (1=pilot, 2=controlor de trafic, 3=navigator de bord). Variabila rang conține poziția a fiecărui subiect sub aspectul reprezentării spațiale, în raport cu toate valorile înregistrate.

Formula de calcul pentru testul Kruskal-Wallis (notat cu H) este următoarea:

$$H = \frac{12}{N*(N+1)} * \sum_{i=1}^{k}\left(\frac{T_i^2}{n_i}\right) - 3*(N+1) \qquad \text{(formula 5.4)}$$

unde:
H este valoarea calculată a testului K-W
N este volumul total al eşantionului
n este volumul grupurilor ($N=n_1+n_2+n_3+...+n_k$)
K este numărul grupurilor independente
T este suma rangurilor care va fi calculată pentru fiecare grup

Înlocuind valorile corespunzătoare exemplului, obținem:

$$H = \frac{12}{13*14} * \left(\frac{41^2}{6} + \frac{23^2}{3} + \frac{27^2}{4}\right) - 3*14 = 0.06593*638{,}74 - 42 = 0{,}1121$$

Valorile distribuției de nul ale lui H urmează forma distribuției chi-pătrat care, ne amintim, are originea în valoarea 0. Cu cât sumele rangurilor pentru cele *k* grupuri sunt mai diferite între ele, cu atât valoarea testului este mai mare şi, potențial, mai aproape de o variație semnificativă. Diferențele mici dintre rangurile grupurilor conduc spre valori ale testului care tind spre 0 şi, implicit, nesemnificative. Valoarea critică a testului se citeşte din tabelul distribuției chi-pătrat pentru df=k-1. Există, totuşi, o excepție, atunci când nici unul din grupurile comparate nu este mai mare de 6, situație în care decizia se ia cu ajutorul unei tabele speciale. În cazul nostru există un grup cu mai mult de cinci subiecți. Ca urmare, scorul critic pentru alfa=0.05 şi 2 grade de libertate este 5.99. Deoarece H calculat este mai mic decât H critic, suntem nevoiți să acceptăm ipoteza de nul şi să concluzionăm că cele trei categorii de subiecți nu sunt diferite sub aspectul capacității de reprezentare spațială.

Copyright © DEPARTAMENT ID 2008

Sarcina de lucru nr. 3.7

Un psiholog trebuie să recomande unui patiser culoarea glazurii pentru un nou tip de prăjitură, având de ales între verde, roşu şi galben. În acest scop alege 18 subiecți, cărora le cere să efectueze o sarcină plictisitoare având la îndemână platouri cu prăjituri glazurate. Subiecții sunt împărțiți în trei grupe, fiecare primind prăjituri de o singură culoare. După un timp, numără câte prăjituri a mâncat fiecare subiect din cele trei grupuri şi construieşte tabelul următor.

Verde	Roşu	Galben
3	3	2
7	4	0
1	5	4
0	6	6
9	4	4
2	6	1

Care este valoarea testului şi care este concluzia cercetării?

Verificați răspunsul corect

Testul Wilcoxon pentru două eşantioane perechi

Dacă avem subiecți evaluați de două ori, pe o scală de interval, iar variabilele nu întrunesc condițiile pentru utilizarea testului *t* al diferențelor pentru eşantioane dependente, se poate apela la testul Wilcoxon. Acesta este un test care, deşi se aplică pe scale de interval/raport, utilizează proceduri de tip neparametric, apelând la diferențele dintre valorile perechi şi la ordonarea lor. Este, din acest punct de vedere, un test de date ordinale.

Exemplu

Un psiholog evaluează frecvența conduitelor agresive după prezentarea unui film care are incluşi stimuli subliminali cu semnificație agresivă. Frecvența conduitelor agresive este măsurată înainte şi după vizionarea filmului. Rezultatele sunt sintetizate în tabelul următor.

Cod subiect	„Înainte”	„După”	„după”-„înainte”	Modulul diferenței	Rangul diferenței	Semnul Diferenței
1	9	8	-1	1	7.5	-
2	14	17	3	3	5.5	+
3	10	17	7	7	2.0	+
4	11	12	1	1	7.5	+
5	12	15	3	3	5.5	+
6	9	13	4	4	3.5	+
7	10	14	4	4	3.5	+
8	14	2	-12	12	1.0	-

Copyright © DEPARTAMENT ID 2008

Coloanele tabelului prezintă etapele procedurii de calcul:

- se calculează diferența dintre variabilele supuse testării
- dacă sunt diferențe nule, se elimină
- se iau în considerare diferențele în valoare absolută
- se construiesc rangurile pentru diferențele în valoare absolută
- se marchează semnul diferențelor pentru fiecare pereche de valori

Din acest punct, calcularea valorilor testului este simplă. Se calculează două valori, T(-) prin însumarea rangurilor diferențelor negative şi T(+) prin însumarea rangurilor diferențelor pozitive. Valoarea cea mai mică dintre ele este rezultatul testului Wilcoxon, al cărui nivel de semnificație se află prin compararea cu valorile critice dintr-o tabelă specială (Anexa 8), în funcție de nivelul alfa ales şi de volumul eşantionului (N). Testul se fundamentează pe ideea că atunci când ipoteza nula este adevărată ar trebui ca suma rangurilor pentru diferențele pozitive să fie egală cu suma rangurilor pentru diferențele negative. Pe măsură ce diferența dintre ele este mai mare, ne îndepărtăm de condiția ipoteza de nul.

Decizia statistică se ia în felul următor:

- Atunci când valoarea calculată este mai mică decât valoarea critică tabelară, ipoteza de nul se respinge iar ipoteza cercetării se confirmă
- Atunci când valoarea calculată este mai mare decât valoarea critică tabelară, ipoteza de nul se acceptă, iar ipoteza cercetării nu se confirmă.

Pentru exemplul nostru, T(+)=28.5 iar T(-)=8.5. Acesta din urmă devine rezultatul testului. Valoarea calculată (8.5) este mai mare decât valoarea critică (4) pentru N=8 şi alfa=0.5 bilateral. Ca urmare, suntem nevoiți să acceptăm ipoteza de nul, considerând neconfirmată ipoteza cercetării. După cum observăm, deşi am calculat testul pentru diferențele negative de rang, am emis concluzia în legătură cu diferențele pozitive, deoarece ele făceau obiectul ipotezei de nul.

Ca şi în cazul testului Mann-Whitney, pentru eşantioane mai mari de 20 distribuția de nul a testului Wilcoxon poate fi aproximată prin distribuția normală. Formula de calcul pentru acest caz este următoarea:

$$Z = \frac{T - [n*(n+1)]/4}{\sqrt{n*(n+1)*(2n+1)/24}} \qquad \text{(formula 5.5)}$$

Exceptând situațiile în care se operează pe eşantioane mici, ca în exemplul de mai sus, calculele sunt destul de laborioase. Din fericire, toate programele avansate de statistică oferă proceduri pentru calcularea automată a acestor teste statistice.

Sarcina de lucru 3.8

Un număr de 10 subiecți şi-au evaluat frica de a vorbi în public pe o scală de la 1 la 10, înainte şi după ce au urmat un tratament psihoterapeutic. Datele cercetării sunt prezentate în tabelul următor:

înainte	9	8	10	9	7	8	6	7	5	9
după	8	5	9	6	8	4	9	7	7	10

(1) Ordonați după rang diferențele, în termeni de mărime absolută (ignorând semnele), dar punând semnele între paranteze, după fiecare diferență.

Copyright © DEPARTAMENT ID 2008

(2) Însumați separat rangurile pentru diferențele pozitive şi negative.
(3) Testați ipoteza că tratamentul psihologic a redus fobia de a vorbi în public.

Verificați răspunsurile corecte

Testul Friedman pentru măsurări repetate

Să presupunem că un psiholog doreşte să studieze relația dintre stilurile de conducere (*laissez-faire*, democratic şi autoritar) asupra nivelului de satisfacție profesională. În acest scop el poate constitui un grup de cercetare pe care să îl supună, în momente succesive, celor trei tipuri de conducere. Un alt model ar putea fi constituirea a trei eşantioane perechi, astfel constituite încât fiecărui subiect dintr-un eşantion să îi corespundă câte un subiect „echivalent" din fiecare dintre celelalte două eşantioane (criteriile de echivalență pot fi: sexul, vârsta, nivelul de inteligență, gardul de motivare, etc.).

Dar, oricare dintre variantele pe care l-ar alege cercetătorul, din punct de vedere statistic el ar obține o structură de date identică: trei serii de evaluări ale satisfacției (variabila dependentă), pentru aceiaşi subiecți (sau perechi de subiecți) corespunzătoare celor trei stiluri de conducere. Dacă variabila dependentă ar fi măsurată pe o scală de interval/raport, testul parametric adecvat este unul care nu a fost inclus în acest volum, „ANOVA pentru măsurări repetate". În lipsa lui şi presupunând că variabila dependentă nu întruneşte condițiile unui test parametric, soluția problemei este testul Friedman pentru date ordinale. Pentru aplicarea lui este suficient ca valorile variabilei dependente să fie ordonate după rang, ca în tabelul alăturat. Facem precizarea că, în acest caz, ordonarea după rang se face la nivelul fiecărui set de evaluări perechi:

	Democratic	***Laissez-faire***	**Autocratic**
1	1	2	3
2	2	1	3
3	1	2	3
4	1	2	3
5	1	2	3
6	2	1	3
N=6	**T_1=8**	**T_2=10**	**T_3=18**

Testul Friedman (F_r) pune în evidență în ce măsură rangurile evaluărilor repetate diferă cu adevărat (statistic semnificativ) unele de altele, după formula:

$$F_r = \frac{12}{N * c * (c+1)} * \sum_{i=1}^{c} T_i^2 - 3 * N * (c+1) \qquad \text{(formula 5.6)}$$

unde:

c este numărul măsurărilor repetate
N este volumul seturilor de evaluări perechi

Copyright © DEPARTAMENT ID 2008

T_i este suma rangurilor corespunzătoare fiecărui moment de măsurare

La fel ca şi în cazul testului H (Kruskal-Wallis), distribuția de nul a testului Friedman urmează forma distribuției chi-pătrat pentru df=c-1.

Introducem valorile cercetării în formulă:

$$F_r = \frac{12}{6*3*4} * \left(8^2 + 10^2 + 18^2\right) - 3*6*4 = 0{,}1666 * 488 - 72 = 9.3$$

Valoarea critică tabelară chi-pătrat pentru df=3-1=2, este 5.99. Valoarea calculată fiind mai mare, se respinge ipoteza de nul şi se consideră confirmată ipoteza cercetării: nivelul satisfacției profesionale variază semnificativ în funcție de stilul de conducere.

Testul Friedman poate fi aplicat şi în cazul a doar două măsurări, situație în care devine similar testului semnului. La fel ca şi celelalte teste pentru date ordinale, el este afectat de existența rangurilor atribuite *ex-aequo*, pentru valori identice. În astfel de cazuri este recomandabilă aplicarea unei corecție formulei de calcul, pe care nu o vom prezenta aici, în speranța că utilizarea programelor specializate va face, oricum, corecțiile necesare.

Sarcina de lucru nr. 3.9

Un neurofiziolog doreşte să verifice dacă există o relație între leziunea cerebrală stângă şi tipul de deficit de memorie de scurtă durată, în trei tipuri de sarcină diferite: cifre, litere, litere şi cifre amestecate. În acest scop, un număr de cinci subiecți cu leziune cerebrală stângă au efectuat teste de memorie distincte, pe şiruri de cifre, litere şi combinații de cifre şi litere. Performanța înregistrată marchează şirul cel mai lung memorat pentru fiecare test în parte.

Datele cercetării:
(valorile semnifică lungimea şirului memorat)

Subiectul	Cifre	Litere	Cifre/Litere
A	6	5	6
B	8	7	5
C	7	7	4
D	8	5	8
E	6	4	7
F	7	6	5

Care este valoarea testului Friedman?
Care este decizia statistică şi ce concluzie trage cercetătorul?

Verificați răspunsurile corecte

Coeficientul de corelație pentru date ordinale (Spearman)

Testele Wilcoxon şi Friedman sunt utilizate pentru a pune în evidență diferențele dintre două sau mai multe eşantioane perechi (situație care, de regulă, se referă la măsurări repetate pe aceiaşi

Copyright © DEPARTAMENT ID 2008

subiecți). Atunci când avem două variabile ordinale şi suntem interesați în evaluarea gradului de asociere între ele, vom utiliza un test similar coeficientului de corelație pentru date de interval care este coeficientul de corelație a rangurilor (Spearman).

Aşa cum ne amintim, coeficientul de corelație Pearson (r) ne dă măsura intensității legăturii dintre două variabile exprimate pe scale de tip interval/raport. Mecanismul de calcul se bazează pe transformarea valorilor ambelor variabile în scoruri z, adică pe convertirea acestora în „distanță standard" față de medie. Pentru datele de tip ordinal, modalitatea de calcul a coeficientului de corelație se bazează pe poziția relativă a unei valori față de celelalte. Coeficientul de corelație a rangurilor Spearman (r_S) are acelaşi domeniu de variație (-1/+1) şi se interpretează în acelaşi mod ca şi coeficientul de corelație pentru date parametrice Pearson.

Exemplu:

Problema cercetării: Într-un studiu cu privire la ameliorarea sistemului de evaluare e personalului, doi instructori urmează un program special de armonizare a evaluării. La sfârşitul programului ei sunt puşi să ierarhizeze personalul unui compartiment de muncă (N=10) din punctul de vedere al performanței profesionale.

Ipoteza cercetării: (pentru test bilateral) Evaluările celor doi instructori vor fi concordante.

Ipoteza de nul: Între evaluările celor doi instructori nu există nici o legătură

Criteriile deciziei statistice:

$\alpha = 0.05$

r_S critic se citeşte într-un tabel special pentru coeficientul Spearman (Anexa 9).

Valoarea se citeşte la intersecția dintre linia corespunzătoare lui N (în acest caz nu se folosesc gradele de libertate) cu coloana corespunzătoare tipului de test (unilateral, bilateral) şi a nivelului α. Înregistrăm r_S critic =0.648

Datele cercetării:

Angajați	**RANG Instructor I**	**RANG Instructor II**	**Diferența (D) (R1-R2)**	**D^2**
A	3	2	1	1
B	1	3	-2	4
C	7	5	2	4
D	6	4	2	4
E	10	10	0	0
F	5	8	-3	9
G	9	7	2	4
H	8	9	-1	1
I	4	6	-2	4
J	2	1	1	1
				$\sum D^2=32$

Formula de calcul pentru coeficientul de corelație a rangurilor Spearman este:

$$r_S = 1 - \frac{6 * \sum D^2}{N * (N^2 - 1)} \qquad \text{(formula 5.7)}$$

În care, prin înlocuirea cu valorile cercetării, obținem:

$$r_S = 1 - \frac{6 * 32}{10 * (100 - 1)} = 1 - \frac{192}{990} = 1 - 0.19 = 0.81$$

60

Copyright © DEPARTAMENT ID 2008

Decizia statistică: r_S calculat $(0.81)^3$ r_S critic (0.684). Ipoteza de nul se respinge.
Concluzia cercetării: Evaluările celor doi instructori sunt semnificativ concordante. Programul de instruire a avut efectul scontat.

Interpretarea coeficientului de corelație Spearman

În principiu, acesta se interpretează la fel ca şi coeficientul Pearson.

$r_S = 0$	Cele două variabile nu variază concomitent, de loc
$0 > r_S > 1$	Cele două variabile tind să crească sau să scadă concomitent, într-o anumită măsură
$r_S = 1.0$	Corelație pozitivă perfectă
$-1 > r_S > 0$	În timp ce o variabilă tinde să crească, cealaltă tinde să descrească
$r_S = -1.0$	Corelație negativă perfectă

Dacă nivelul de semnificație (*p*) este mai mare decât 0.05, coeficientul de corelație va fi considerat nesemnificativ (are şanse prea mari să rezulte din jocul întâmplării). Aceasta nu înseamnă că nu există o corelație între cele două variabile ci doar că datele noastre nu au putut să o pună în evidență.

Calcularea coeficientului de determinare (r^2) în cazul corelației Spearman nu este recomandabilă, deşi există autori care o acceptă.

Când se utilizează coeficientul de corelație Spearman:

- Atunci când ambele variabile sunt de tip ordinal
- Atunci când una dintre variabile este de tip ordinal şi cealaltă este de tip interval/raport. În acest caz, variabila interval/raport se transformă mai întâi în valori de ordine de rang
- Atunci când ambele variabile sunt de tip interval/raport dar una sau ambele, prezintă valori extreme. În acest caz, prin transformarea în ordine de rang a celor două distribuții, valorile extreme sunt anihilate, ele urmând să participe la corelație prin simpla poziție în distribuție şi nu prin nivelul lor absolut.

Un test alternativ pentru asocierea variabilelor ordinale este coeficientul de corelație a rangurilor *Kendall tau*. La fel ca şi coeficientul Spearman, Kendal tau ia valori între -1 şi +1. Similaritățile se opresc, însă, aici deoarece coeficientul Kendall se calculează pe o cale diferită şi se fundamentează pe o estimare a parametrului populației. Aceasta estimare se calculează ca probabilitatea concordanței minus probabilitatea discordanței dintre rangurile perechi. Nu vom analiza în amănunt procedura de calcul, dar vom prezenta modul de obținere a coeficientului Kendall cu ajutorul programului SPSS în secțiunea următoare a acestui volum.

Ambii coeficienți sunt larg utilizați în studiile statistice, făcând, în acelaşi timp, şi obiectul unor dispute între statisticieni. Adesea coeficientul Kendall este considerat mai adecvat datorită faptului că distribuția acestuia se apropie de forma normală începând de la volume mai mici ale eşantioanelor. Chiar dacă în calcule pe aceleaşi date cu cei doi coeficienți obțin valori diferite, decizia statistică nu este, de obicei diferită.

Sarcina de lucru 3.10

Copyright © DEPARTAMENT ID 2008

Un psiholog doreşte să verifice relația dintre inteligența verbală şi cea non verbală. În acest scop, un eşantion de 10 subiecți este supus unei testări cu două teste care vizează cele două categorii de inteligență, exprimate în unități QI. Din cauza volumului foarte redus al eşantionului, decide să aplice testul de corelație a rangurilor Spearman, în locul testului Pearson.
Datele cercetării sunt următoarele:

Subiect	Qi verbal	QI nonverbal
A	110	105
B	100	108
C	120	110
D	90	95
E	108	105
F	115	125
G	122	118
H	110	116
I	127	118
J	118	126

Calculați testul de corelație Spearman şi enunțați decizia statistică şi concluzia cercetării
Verificați rezultatul corect la pagina 38

Rezumatul unității de învățare

- Pentru a testa dacă două populații diferă între ele, pe o variabilă continuă, fără a avea posibilitatea de măsurare exactă a acesteia (pe scală de interval sau raport), pot fi selecționate două eşantioane care sunt puse împreună, după care valorile sunt ordonate pe baza rangurilor de mărime. Testarea diferenței se face prin însumarea separată a rangurilor valorilor celor două eşantioane.
- Dacă eşantioanele nu diferă, suma rangurilor va fi apropiată sau egală, dacă diferă, semnificația diferenței dintre ranguri este probată cu testul Mann-Whitney U.
- Testul Mann-Whitney U este utilizat ori de câte ori o variabilă nu poate fi măsurată precis, dar se poate determina ordinea valorilor. De asemenea, el se utilizează atunci când cele două variabile sunt măsurate pe scale cantitative dar prezintă valori aberante legitime.
- Testul Mann-Whitney este echivalentul pentru date ordinale a testului diferenței dintre medii pentru eşantioane independente (compuse din subiecți diferiți).
- Testul Kruskal-Wallis este o extensie a testului Mann-Whitney şi se utilizează atunci când avem de comparat rangurile a mai mult de două eşantioane independente. Din acest punct de vedere, Testul Kruskal.Wallis este echivalentul pentru date ordinale al analizei de varianță unifactoriale (ANOVA).
- Dacă datele sunt recoltate de la aceiaşi subiecți în două condiții de cercetare diferite, testarea diferenței dintre ranguri se face cu testul Wilcoxon. Acesta este echivalentul testului t pentru diferența dintre mediile a două eşantioane dependente (compuse din aceiaşi subiecți).
- În acest caz, toate diferențele sunt ordonate după mărime, ignorând semnul lor, suma rangurilor fiind făcută separat, pentru diferențele pozitive şi negative. O diferență mare dintre aceste două sume este dovada unei diferențe între cele două eşantioane comparate.

Copyright © DEPARTAMENT ID 2008

- Testul Friedman, este un test care extinde cazul testul Wilcoxon la mai mult de două eşantioane dependente. El este echivalentul analizei de varianță pentru măsurători repetate (test care nu a fost tratat în acest volum).
- Atunci când avem două variabile măsurate pentru aceiaşi subiecți, ale căror valori se exprimă pe scală ordinală, şi dorim să testăm gradul de asociere dintre acestea, în locul testului de corelație Pearson se utilizează corelația Spearman, pentru date ordinale.
- Corelația Spearman este recomandabilă şi atunci când variabilele sunt măsurate pe scale de interval sau de raport, dar distribuția uneia, sau a ambelor, se îndepărtează grav de la forma normală.
- Coeficientul de corelație Spearman se defineşte în aceeaşi plajă de valori ca şi corelația Pearson (±1) şi se interpretează în mod similar.
- Testele bazate pentru variabile ordinale sunt independente de forma distribuției variabilelor.

Răspunsuri la sarcinile de lucru

Răspunsul la sarcina de lucru 3.6

Două grupuri de câte 5 subiecți au fost supuse unui experiment care vizează reacția fiziologică la prezentarea unui stimul şocant. Unul din grupuri a primit un instructaj care sugera natura şocantă a stimulului în timp ce celuilalt grup i s-a spus că vor primi stimuli agreabili. Datele de mai jos prezintă valoarea pulsului măsurată în primele 10 secunde de la prezentarea stimulului:

Grupul cu instructaj liniştitor:	Rang
105	7
130	2
145	1
125	3
115	5
Suma rangurilor	**18**
Grupul cu instructaj pregătitor corect	
120	4
80	10
90	9
110	6
95	8
Suma rangurilor	**37**

$$U_A = 5*5 + \frac{5*(5+1)}{2} - 18 = 22$$

$$U_B = 5*5 + \frac{5*(5+1)}{2} - 37 = 3$$

Valoarea U_B este valoarea testului şi este egală cu valoarea tabelară (tabela 3.3) pentru n_A=5 şi n_B=5, alfa=0.05. În acest caz se respinge ipoteza de nul şi se acceptă ipoteza cercetării. Anticiparea naturii stimulului reduce reacția fiziologică de surpriză.

Răspunsul la sarcina de lucru 3.7

Copyright © DEPARTAMENT ID 2008

(1)

după	4	5	6	9	7	8	9	8	10	7
înainte	8	8	9	6	5	9	10	7	9	7
	4(-)	3(-)	3(-)	3(+)	2(+)	1(-)	1(-)	1(+)	1(+)	0
rang	***1***	***3***	***3***	***3***	***5***	***7.5***	***7.5***	***7.5***	***7.5***	***10***

(2) Suma diferențelor pozitive=23; Suma diferențelor negative=22
(3) Valoarea testului Wilcoxon este T(-)=22, mai mare decât valoarea critică tabelară 8 (pentru alfa=0.05 şi N=10, din tabelul 3.3, în anexe). Se respinge ipoteza de nul şi se admite ipoteza cercetării. Terapia are efect (chiar dacă diferența nu este una foarte mare).

Răspunsul la sarcina de evaluare 3.8

	prăjituri mâncate	**rang**	
verde	9	1	
verde	7	2	
Roşu	6	4	
Roşu	6	4	
galben Roşu	6 5	4 6	
Roşu	4	8.5	
Roşu	4	8.5	
galben galben verde	4 4 3	8.5 8.5 11.5	
Roşu	3	11.5	
verde	2	13.5	
galben verde	2 1	13.5 15.5	
galben verde	1 0	15.5 17.5	
galben	0	17.5	

Suma rangurilor pe culori:
Roşu=31
Verde=61
Galben=67.5

- H critic=5.99, citit în tabela chi-pătrat pentru alfa=0.05 şi 2 grade de libertate (numarul grupurilor minus 1)
- Calculăm testul Kruskal Wallis:

$$H = \frac{12}{18*(18-1)}*\left(\frac{31^2}{6}+\frac{61^2}{6}+\frac{67.5^2}{6}\right)-3*(18-1)=9.04$$

- Hcalculat (9.04) este mai mare decât Hcritic (5.99).
- Decizia statistică: respingem ipoteza de nul.

Copyright © DEPARTAMENT ID 2008

- Se admite că preferința pentru prăjitură este influențată de culoare acesteia. Cele mai preferate sunt cele de culoare roşie (plasate pe ranguri superioare), urmate de culoarea verde şi galbenă, care se află la mică distanța una de cealaltă.

Răspunsul la sarcina de lucru 3.9

Subiectul	Cifre	***rang***	Litere	***rang***	Cifre/Litere	***rang***
A	6	***1.5***	5	***2***	6	***1.5***
B	8	***1***	7	***2***	5	***3***
C	7	***1.5***	7	***1.5***	4	***3***
D	8	***1.5***	5	***3***	8	***1.5***
E	6	***2***	4	***3***	7	***1***
F	7	***1***	6	***2***	5	***1***
Suma rangurilor		***8.5***		***13.5***		***11***

Valoarea critică citită în tabelul chi-pătrat, pentru df=3-1=2 şi alfa=0.05, este 5.99

Calculăm testul Friedman:

$$Fr = \frac{12}{6*3*(3+1)} * \left(8.5^2 + 13.5^2 + 11^2\right) - 3*6*(3+1) = 48.16$$

Fr calculat este mai mare decât Fr critic. Ca urmare, se respinge ipoteza de nul şi se acceptă ipoteza cercetării. Leziunea cerebrală stângă are un efect asupra conținutului memorat. Cea mai afectată este memorarea literelor (obține ranguri superioare) iar cea mai puțin afectată este memorarea cifrelor (obține cele mai mici ranguri).

Răspunsul la sarcina de lucru 3.10

Subiect	**Qi verbal**	**Rang v.**	**QI nonverbal**	**Rang nv.**	**$(v\text{-}nv)^2$**
A	110	6.5	105	8.5	6.25
B	100	9	108	4	25
C	120	3	110	7	64
D	90	10	95	10	0
E	108	8	105	8.5	0.25
F	115	5	125	2	9
G	122	2	118	4	4
H	110	6.5	116	6	0.25
I	127	1	118	4	9
J	118	4	126	1	9

$\Sigma(v\text{-}nv)^2=126.75$

Valoarea critică a testului se citeşte în tabela 3.5 din anexe, pentru alfa 0.05 bilateral şi N=10:
Rs critic=0.64

Calculăm Rs

$$r_s = 1 - \frac{6*126.75}{10*(100-1)} = 0.28$$

Copyright © DEPARTAMENT ID 2008

Decizia statistică: Se acceptă ipoteza de nul. Rs calculat (0.28) este mai mic decât rs critic (0.64). Performanța la cele doua teste de inteligență nu corelează semnificativ.

Lucrarea de evaluare nr. 3.2 şi modul de cotare

Lucrarea de evaluare va fi publicată pe portal. Termenul de trimitere este preziua tutorialului. După această dată lucrările trimise nu vor mai fi acceptate.

Bibliografie minimală

Bibliografia de bază

- Marian Popa, (2008), Statistică pentru psihologie. Teorie □i aplica□ii SPSS, Polirom
- Pagina web a cursului, la adresa: www.*mpopa.ro*

Bibliografie suplimentară

- Clocotici, V., & Stan, A. (2000). Statistica aplicata in psihologie. Iasi: Polirom, (selectiv)
- Rotaru, T. (coord.). (1999). Metode statistice aplicate in stiintele sociale. Iasi: Polirom. (selectiv)
- Vasilescu, I. P. (1992). Statistica informatizata pentru stiinte despre om (Vol. 1-2). Bucuresti: Editura militara., (selectiv)

Copyright © DEPARTAMENT ID 2008

ANEXE – TABELE STATISTICE

Anexa 6. Valorile critice pentru distribuția *chi-pătrat*

(sursa:http://www.psychology.ilstu.edu/psy138/chi)

	p		
df	***.05***	***.025***	***.01***
1	3.84	5.02	6.64
2	5.99	7.38	9.21
3	7.81	9.35	11.34
4	9.49	11.14	13.28
5	11.07	12.83	15.09
6	12.59	14.45	16.81
7	14.07	16.01	18.48
8	15.51	17.53	20.09
9	16.92	19.02	21.67
10	18.31	20.48	23.21
11	19.68	21.92	24.72
12	21.03	23.34	26.22
13	22.36	24.74	27.69
14	23.68	26.11	29.14
15	25.00	27.49	30.58
16	26.30	28.85	32.00
17	27.59	30.19	33.41
18	28.87	31.53	34.80
19	30.14	32.85	36.19
20	31.41	34.17	37.57
21	32.67	35.48	38.93
22	33.92	36.78	40.29
23	35.17	38.08	41.64
24	36.42	39.36	42.98
25	37.65	40.65	44.31
26	38.88	41.92	45.64
27	40.11	43.19	46.96
28	41.34	44.46	48.28
29	42.56	45.72	49.59
30	43.77	46.98	50.89
40	55.76	59.34	63.69
50	67.50	71.42	76.15
60	79.08	83.29	88.38
70	90.53	95.02	100.42
80	101.88	106.63	100.43
90	113.15	118.14	124.12
100	124.34	129.56	135.81

Copyright © DEPARTAMENT ID 2008

Anexa 7. Tabelul valorilor critice pentru testul Mann-Whitney U

(sursa: Clocotici V., Stan A., 2000, Statistică aplicată în psihologie, Polirom)

n_A/n_B	α	5	6	8	10	12	14	16	18	20
3	***0.05***	0	1	2	3	4	5	6	7	8
	0.01	-	-	-	0	1	1	2	2	3
4	***0.05***	***1***	***2***	***4***	***5***	***7***	***9***	***11***	***12***	***14***
	0.01	-	0	1	2	3	4	5	6	8
5	***0.05***	***2***	***3***	***6***	***8***	***11***	***13***	***15***	***18***	***20***
	0.01	0	1	2	4	6	7	9	11	13
6	***0.05***	3	5	8	11	14	17	21	24	27
	0.01	1	2	4	6	9	11	13	16	18
8	***0.05***	6	8	13	17	22	26	31	36	41
	0.01	2	4	7	11	15	18	22	26	30
10	***0.05***	8	11	17	23	29	36	42	48	55
	0.01	4	6	11	16	21	26	31	37	42
12	***0.05***	11	14	22	29	37	45	53	61	69
	0.01	6	9	15	21	27	34	41	47	54
14	***0.05***	13	17	26	36	45	55	64	74	83
	0.01	7	11	18	26	34	42	50	58	67
16	***0.05***	15	21	31	42	53	64	75	86	98
	0.01	9	13	22	31	41	50	60	70	79
18	***0.05***	18	24	36	48	61	74	86	99	112
	0.01	11	16	26	37	47	58	70	81	92
20	***0.05***	20	27	41	55	69	83	98	112	127
	0.01	13	18	30	42	54	67	79	92	105

Copyright © DEPARTAMENT ID 2008

Anexa 8. Valorile critice pentru testul Wilcoxon

(sursa : Clocotici V., Stan A., 2000, Statistică aplicată în psihologie, Polirom)

N	Nivel de seminficație pentru test unilateral		
	0.025	**0.01**	**0.005**
	Nivel de seminficație pentru test bilateral		
	0.05	**0.02**	**0.01**
6	0	-	-
7	2	0	-
8	4	2	0
9	6	3	2
10	8	5	3
11	11	7	5
12	14	10	7
13	17	13	10
14	21	16	13
15	25	20	16
16	30	24	20
17	35	28	23
18	40	33	28
19	46	38	32
20	52	43	38
21	59	49	43
22	66	56	49
23	73	62	55
24	81	69	61
25	89	77	68

Copyright © DEPARTAMENT ID 2008

Anexa 9. Valorile critice pentru testul de corelație a rangurilor (Spearman)
(sursa: http://www.netnam.vn/unescocourse/index.htm)

N	test unilateral			
	alfa=0.05	alfa=0.025	alfa=0.01	alfa=0.005
	test bilateral			
	alfa=0.10	alfa=0.05	alfa=0.02	alfa=0.01
5	0,900			
6	0,829	0,886	0,943	
7	0,714	0,786	0,893	
8	0,643	0,738	0,833	0,881
9	0,600	0,683	0,783	0,833
10	0,564	0,648	0,745	0,794
11	0,523	0,623	0,736	0,818
12	0,497	0,591	0,703	0,780
13	0,475	0,566	0,673	0,745
14	0,457	0,545	0,646	0,716
15	0,441	0,525	0,623	0,689
16	0,425	0,507	0,601	0,666
17	0,412	0,490	0,582	0,645
18	0,399	0,476	0,564	0,625
19	0,388	0,462	0,549	0,608
20	0,377	0,450	0,534	0,591
21	0,368	0,438	0,521	0,576
22	0,359	0,428	0,508	0,562
23	0,351	0,418	0,496	0,549
24	0,343	0,409	0,485	0,537
25	0,336	0,400	0,475	0,526
26	0,329	0,392	0,465	0,515
27	0,323	0,385	0,456	0,505
28	0,317	0,377	0,448	0,496
29	0,311	0,370	0,440	0,487
30	0,305	0,364	0,432	0,478

Copyright © DEPARTAMENT ID 2008